DOCUMENT RESUME

ED 041 631 PS 003 298

AUTHOR	Stern, Carolyn
TITLE	Maximizing the Value of Evaluation for the Head Start Teacher. Final Report.
INSTITUTION	California Univ., Los Angeles. Center for Head Start Evaluation and Research.
SPONS AGENCY	Office of Economic Opportunity, Washington, D.C.
REPORT NO	OEO-4117
PUB DATE	31 Aug 69
NOTE	188p.
EDRS PRICE	EDRS Price MF-$0.75 HC-$9.50
DESCRIPTORS	Behavior Rating Scales, *Classroom Observation Techniques, *Evaluation Techniques, *Feedback, Measurement Instruments, Teacher Attitudes, *Teacher Evaluation
IDENTIFIERS	*Project Head Start

ABSTRACT

The purpose of this study was to find out whether systematic evaluation feedback to teachers would result in: (1) development of more favorable attitudes toward evaluation, as measured by Teacher Attitudes Toward Evaluation (TATE), and increased teacher use of a variety of curricula and materials, (2) greater correspondence between teachers' expectations and children's performance, as measured by Teachers Expectations for Achievement of Children in Head Start (TEACH), and (3) greater gains on cognitive and affective measures for children in feedback classes. 183 children were tested on a variety of measures. Data were also collected from classroom observations and parent interviews. Of the 24 classes in the study, 10 classes (20 teachers) were in the feedback group, which attended monthly meetings where they were informed about children's test results, test instruments, and classroom observations. Teachers also viewed video tapes of their own classes. Although children showed consistent gains, teachers given feedback were not measurably more successful in remediating deficiencies than those not receiving feedback. With reference to teacher attitudes, the intervention was more effective, as shown by TEACH and TATE scores and by increased understanding and rapport between teachers and evaluators. (NH)

F-OEO - 4117
PS

ED041631

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

MAXIMIZING THE VALUE OF EVALUATION FOR THE HEAD START TEACHER

FINAL REPORT

UCLA Head Start Research and Evaluation Center

Research Projects in Early Childhood Learning

Carolyn Stern, Director

University of California, Los Angeles
Graduate School of Education

The work reported herein was carried out under a contract with the United States Office of Economic Opportunity, No. 4117.

September 1, 1968 to August 31, 1969

PS003298

ACKNOWLEDGEMENTS

During the course of this research intervention study, many people made a wide variety of contributions. The staff of the UCLA Head Start Evaluation and Research Center are listed in the Personnel section of this report. To thank individually each member who made important contributions to the conduct of the Feedback Intervention would be tantamount to repeating almost all the names in that list. However, special thanks are due to our sensitive and perceptive Evaluation Coordinator, Liz Raymer and to Willa Gupta, who struggled so valiantly to make sense of the mountains of data which were generated. Appreciation for the level-headed and soothing intervention of Dr. Joseph Edwards in a number of volatile crisis situations was only belatedly expressed when he was appointed Assistant Director of the Center in July, 1969.

Yet, in one sense, all of the Center staff were doing a job; doing it superbly well, but still it was what they were being paid to do. It is to all the Head Start program personnel, who gave so freely of their time and efforts, that the successful completion of this project is ultimately due. Without the cooperation of Mr. Charles Taylor and Mr. Charles Manfred of the Los Angeles Economic and Youth Opportunities Agency, and Mrs. Grace Blagdon and Mr. Gary Alexander of the Compton-Willowbrook-Enterprise Community Action Agency, the project would never have gotten off the ground. Mrs. Nancy Mattias of the Foundation for Early Childhood Education, Mr. Leo Henderson of the Los Angeles Urban League, and Mr. Brice Taylor of the Kedren Community Health Center, were always available and ready to smooth out any problems that arose. And Mr. Robert Wright of the Los Angeles Council of Churches offered several of his sites for practice and training of our staff in the administration of some of the new instruments and observation procedures.

Even the most enthusiastic and supportive top-level endorsement of a project can be subtly sabotaged if the operating personnel are disaffected and opposed to its purposes. With 24 Head Teachers and 24 Assistant Teachers, there was indeed a wide range of involvement; undoubtedly there were individuals among this group of 48 who were hostile and unreceptive to the unavoidable interference engendered by the evaluation activities. It is surprising that there were actually so few who were not eventually won over to a more positive attitude--to these we offer our regrets and apologies. But to the large number of teachers who welcomed us and helped us complete our task, we are most sincerely grateful.

Finally, we are indebted to the Office of Economic Opportunity not only for the funding which made all this possible, but also for the warm cooperation of the national staff, and particularly for the guidance, understanding, and patience of the Director of Research and Evaluation, Dr. John McDavid, and his two right arms, Dr. Lois-ellin Datta and Dr. Edith Grotberg, Coordinators for Evaluation and Research, respectively.

TABLE OF CONTENTS

Page

Acknowledgments

Abstract vi

Part I. Organization, Personnel, and Description of Programs. 1

Organization Chart. 2

Head Start Evaluation and Intervention Program. 3

Head Start Language 5

Follow-Up Study 7

Personnel List 8

Part II. Head Start Evaluation and Intervention Program 12

Maximizing the Value of Evaluation for the Head Start Teacher . . 13

Introduction 13

Hypotheses 14

General Description of Delegate Agencies and Head Start Sites. 15

Subjects 23

Criterion Measures 28

Individual Child. 29

Educational Environments. 32

Family Relationships. 36

Procedure. 37

Meetings and Feedback Sessions with Head Start Teachers. . . 40

General Procedure 40

Orientation Meetings. 40

Feedback Meetings 41

Description of National Head Start Common Core Measures. . . 55

The Data Aide Program 59

Results and Discussion 63

Appendixes to Part II 110

A. Forms

UCLA Feedback Record Sheet. 111

Parent Permission Slips (English and Spanish) 112

TABLE OF CONTENTS (cont'd.)

Page

B. Instructions to Field Staff 114
C. Quality Control Procedures. 118
D. Procedure for Analysis of OSCI Data 120
E. Interim Report After Site Visit 123

Part III. OSCI Training Report 127
Procedure 128
Results and Discussion. 129
Summary 131
Appendixes to Part III 132
A. Schedule of 1968 OSCI Training Sessions 133
B. OSCI Training Procedures Used by UCLA OSCI Trainers, Fall 1969 135
C. Calculation of Inter-Rater Observer Reliability 136
D. OSCI Training Exercise. 138
E. OSCI Coding Posttest. 142
F. Coding Test for OSCI Observers I. 143
G. Coding Test for OSCI Observers II 147

Part IV. Reports of Research Associated with the Evaluation Study. . 151
A. The Development of an Instrument to Measure Teacher Attitudes Toward Evaluation (TATE). 152
B. Teachers Expectations for Achievement of Children in Head Start (TEACH) 164
C. Teacher Ratings of Behavioral Objectives as Related to Performance of Children on Specific Tasks 174

LIST OF TABLES

Page

1. Experimental Design and Allocation of Classes to Treatments . . . 22
2a. Race, Sex, and Age of Sample Population, by Class 25
2b. Race, Sex, and Age of Sample Population, by Treatment 26
3. Percent Attendance and Attrition by Class, Agency, and Treatment 27
4a. Pre, Post and Gain Scores on Binet Vocabulary Subtest (with Means for Subjects Dropped Before Posttesting) 65
4b. Pre, Post, and Gain Scores on Binet Verbal Subtest (with Means for Subjects Dropped Before Posttesting) 66
4c. Pre, Post, and Gain Scores on Binet Performance Subtest (with Means for Subjects Dropped Before Posttesting) 67
5. Means, Standard Deviations, Gain Scores, and Significance of Difference on Pre-Post Stanford Binet by Classes 68
6. Means, Standard Deviations, Gain Scores, and Significance of Difference on Pre-Post Caldwell by Classes 69
7. Means, Standard Deviations, Gain Scores, and Significance of Difference on Pre-Post WPPSI by Classes 70
8. Means, Standard Deviations, Gain Scores, and Significance of Difference on Pre-Post Gumpgookies by Classes 71
9. Means, Standard Deviations, Gain Scores, and Significance of Difference on Pre-Post Stanford Binet by Sites, Agencies, and Treatments 72
10. Means, Standard Deviations, Gain Scores, and Significance of Difference on Pre-Post Caldwell by Sites, Agencies, and Treatments 73
11. Means, Standard Deviations, Gain Scores, and Significance of Difference on Pre-Post WPPSI by Sites, Agencies, and Treatments . 74
12. Means, Standard Deviations, Gain Scores, and Significance of Difference on Pre-Post Gumpgookies by Sites, Agencies, and Treatments 75
13. One-Way Analyses of Variance on Gain Scores for Individual Measures (over 24 classes) 77
14. Pre-Post Analyses of Covariance for Individual Measures (over 24 classes) 77
15. Testing Conditions 78
16. Number of People Attending Each Meeting 80
17. Characteristics of Teaching Staff 82
18. Post Observation of Teacher with D Indicating Direction of Pre-Post Change 85

LIST OF TABLES (cont'd.)

Page

19. Means and Standard Deviations for Experimental and Control Groups with Description of P.O.T. Items Used in Analysis 87
20a. Hand Analysis of OSCI Data for Classroom Observations by Agency and Treatment 89
20b. Hand Analysis of OSCI Data for Teacher Input by Agency and Treatment 90
20c. Hand Analysis of OSCI Data for Involvement by Agency and Treatment 92
21. Variety of Materials Used by Categories 94
22. Observer's Report on the Class 95
23. Observer's Report on the Class: Summary of Open-Ended Items . . 96
24. Class Facilities (by Classes) 102
25. Class Facilities (Means and Standard Deviations for Experimental and Control Groups) 103
26. Means and Standard Deviations for All Variables for All Children 105
27. Significant Correlations for Individual Child Variables 107
28. Correlation Matrix for All Class Variables 109

Part III. OSCI Training Report

1. Reliability of Observers 130

Part IV. Reports of Research Associated with the Evaluation Study

A. The Development of an Instrument to Measure Teacher Attitudes Toward Evaluation (TATE)

1. Mean Pretest and Posttest Scores and Rank Order of Items on Attitude (A) and Belief (B) Scales of Tate 160
2. T-Test Values for Differences between Mean Scores on Pretest and Posttest for Attitude and Belief Scales (by Treatment Group) 162

B. Teachers Expectations for Achievement of Children in Head Start (TEACH)

1. Teachers Expectations for Achievement of Children in Head Start (TEACH) 169
2. Mean Pre- and Posttest Scores and Differences between Importance and Expectation Ratings by Factors and Items . 170
3. Rank Order of Means for 7 Factors on the Teachers Expectations for Achievement of Children in Head Start (TEACH) Scale (by Treatment Groups) 173

PART I

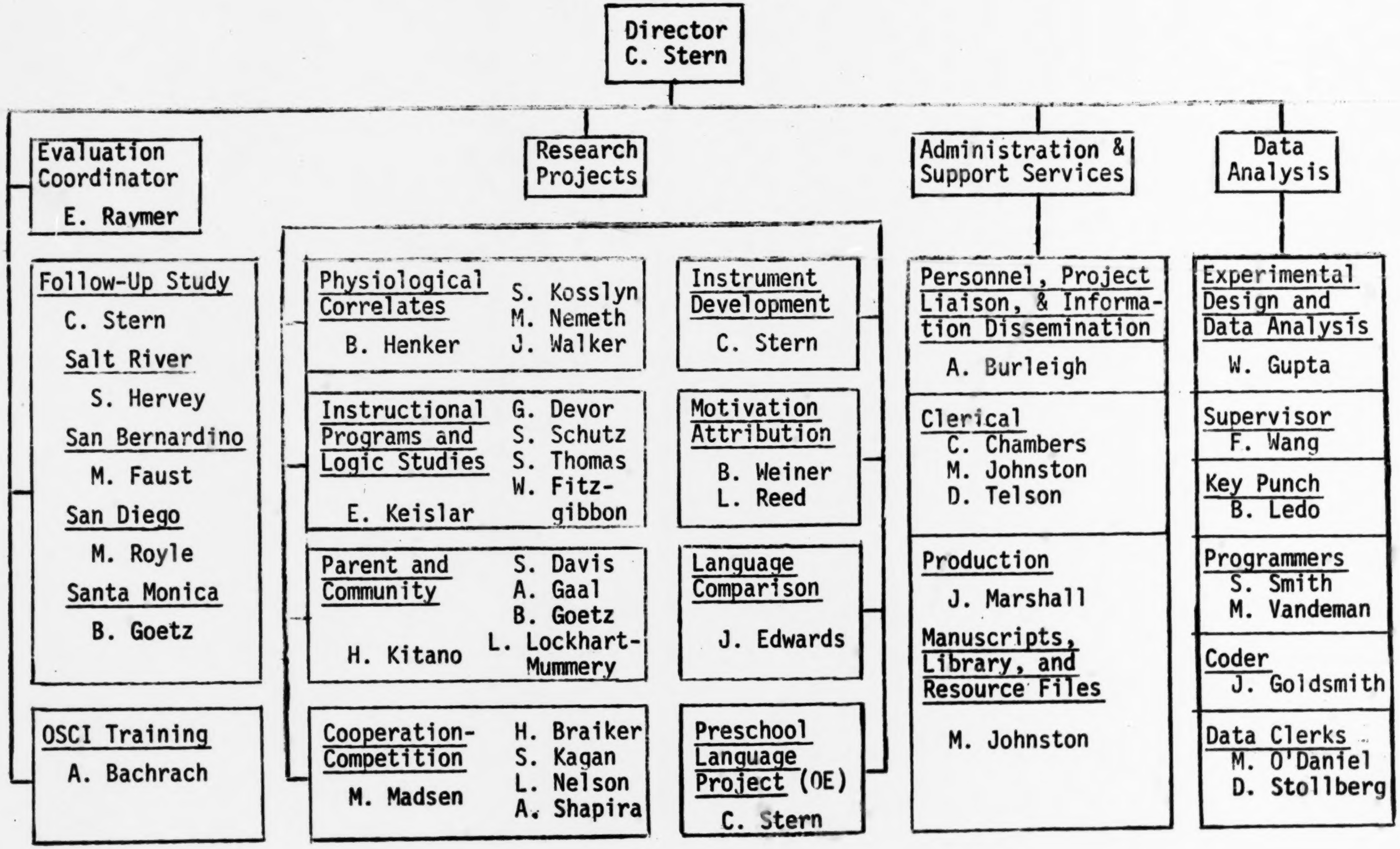
ORGANIZATION, PERSONNEL, AND DESCRIPTION OF PROGRAMS

This section includes the Organization Chart which describes the staffing for both the Head Start Evaluation and Research Center, funded by the Office of Economic Opportunity, and the Preschool Language Project, funded by the United States Office of Education. Many of the staff were employed on both projects and several performed various types of services. For instance, in the evaluation component, the Data Aides were trained to administer the less demanding tests such as the Peabody, Gumpgookies, and Preschool Inventory, as well as the Parent Interviews. The OSCI Observers did parent and teacher interviews and some of the individual child tests. Two of the teachers in the Head Start Language Study also served as teachers for the Preschool Language Project. Administrative, secretarial, and support personnel worked on either project as their services were needed.

On the pages following the Organization Chart for the total Center are the Organization Charts for the Evaluation and Intervention component and the Language Study component, with a brief description of the rationale for each study. A page describing the design of the Follow-Up is also included. These materials are followed by a briefly annotated alphabetical listing of personnel.

RESEARCH PROJECTS IN EARLY CHILDHOOD LEARNING

Organization Chart


Director
C. Stern
Evaluation Coordinator
E. Raymer
Follow-Up Study
C. Stern
Salt River
S. Hervey
San Bernardino
M. Faust
San Diego
M. Royle
Santa Monica
B. Goetz
OSCI Training
A. Bachrach
Research Projects
Physiological Correlates
B. Henker
S. Kosslyn
M. Nemeth
J. Walker
Instructional Programs and Logic Studies
E. Keislar
G. Devor
S. Schutz
S. Thomas
W. Fitz-gibbon
Parent and Community
H. Kitano
S. Davis
A. Gaal
B. Goetz
L. Lockhart-Mummery
Cooperation-Competition
M. Madsen
H. Braiker
S. Kagan
L. Nelson
A. Shapira
Instrument Development
C. Stern
Motivation Attribution
B. Weiner
L. Reed
Language Comparison
J. Edwards
Preschool Language Project (OE)
C. Stern
Administration & Support Services
Personnel, Project Liaison, & Information Dissemination
A. Burleigh
Clerical
C. Chambers
M. Johnston
D. Telson
Production
J. Marshall
Manuscripts, Library, and Resource Files
M. Johnston
Data Analysis
Experimental Design and Data Analysis
W. Gupta
Supervisor
F. Wang
Key Punch
B. Ledo
Programmers
S. Smith
M. Vandeman
Coder
J. Goldsmith
Data Clerks
M. O'Daniel
D. Stollberg

HEAD START EVALUATION AND INTERVENTION PROGRAM

The major purpose of the Head Start Evaluation Program is to provide the Office of Economic Opportunity with objective data regarding changes in children resulting from their Head Start experience. In addition to individual child measures, there is extensive instrumentation for gathering information about program and curriculum variables, classrooms, teachers, classroom facilities, parents, nutrition, and medical services.

At each of the national Head Start Evaluation and Research Centers, the evaluation program included both the administration of the national battery of test instruments and an experimental intervention. The UCLA intervention consisted of a feedback study designed to test the hypotheses that providing systematic information to Head Start teachers about test results will 1) develop more favorable attitudes toward evaluation and that these teachers will improve and enrich their programs as a result of the data they receive; 2) children in these feedback classes will demonstrate greater gains on cognitive and affective measures than children in classes not receiving feedback.

There were 24 Head Start classes in 10 centers involved in the evaluation program. Ten classes (20 teachers) were in the feedback group. Meetings with these teachers were held monthly and information about test results, instruments, and observations were discussed. Teachers also viewed video tapes of their own classes in operation. One control group of eight classes (16 teachers) were given the standard evaluation package but no feedback beyond an initial orientation meeting. A second control group consisting of six classes (12 teachers) received only an abbreviated battery of individual child measures. This group was included to control for the effect of frequent presence of observers and testers in the classroom. The schedule of activities for all groups is presented on page 38.

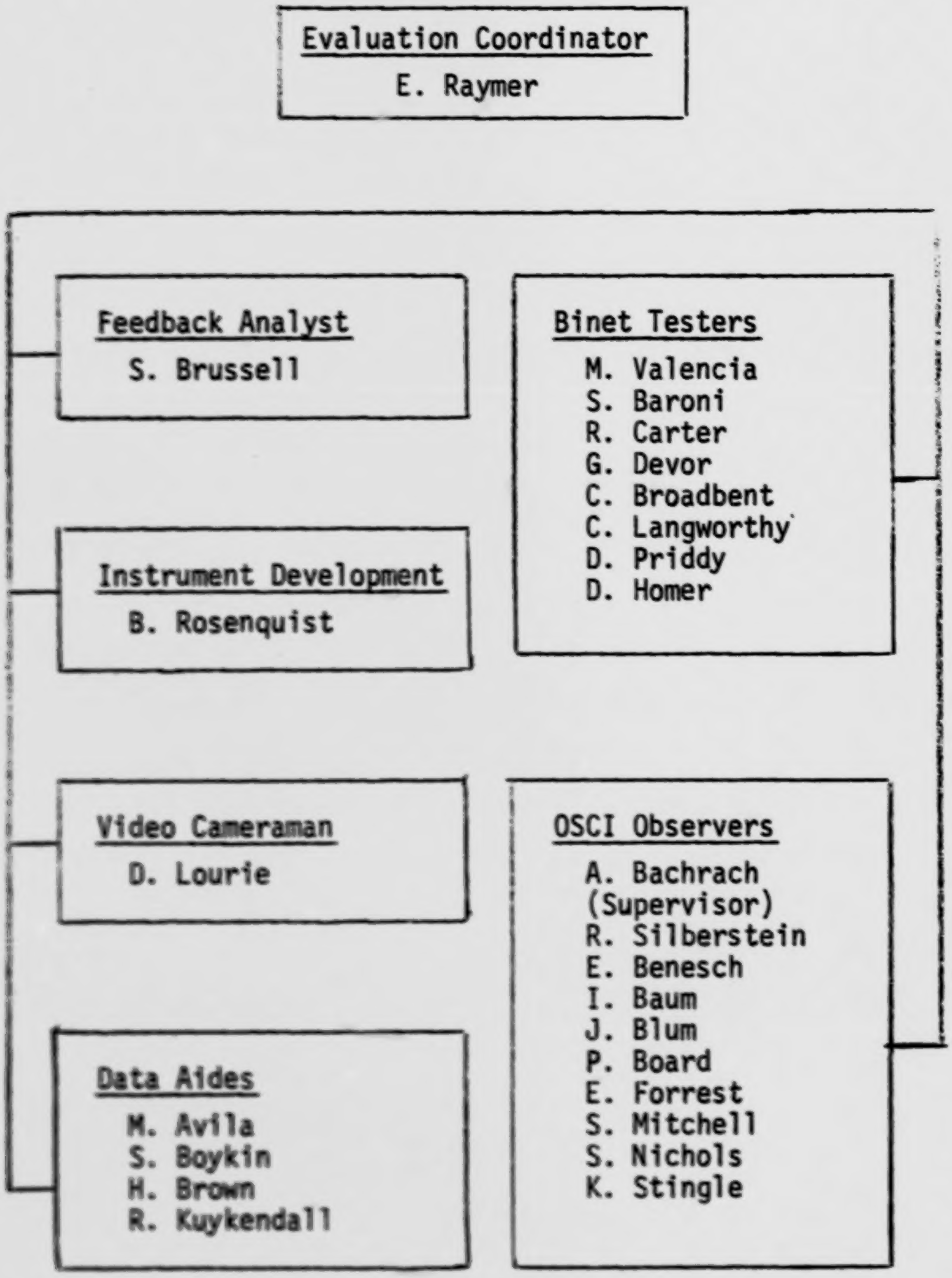
Evaluation Coordinator

The Evaluation Coordinator arranged for the evaluation sites, planned and supervised the data collection schedules, and supervised the collection of data and classroom observations. In addition, she supervised the intervention feedback program and made contact with the Head Start agencies involved.

Evaluation and Intervention Staff

The Feedback Analyst planned and conducted feedback meetings with all the teachers involved in the feedback groups. Data Aides were responsible for data gathering at specific centers and for maintaining good relations between the various Head Start sites and the Evaluation Center. Observers were brought in at the specified observation periods. Quality control was obtained as described on the quality control memorandum.

EVALUATION AND INTERVENTION



Evaluation Coordinator
E. Raymer

Feedback Analyst
S. Brussell

Instrument Development
B. Rosenquist

Video Cameraman
D. Lourie

Data Aides
M. Avila
S. Boykin
H. Brown
R. Kuykendall

Binet Testers
M. Valencia
S. Baroni
R. Carter
G. Devor
C. Broadbent
C. Langworthy
D. Priddy
D. Homer

OSCI Observers
A. Bachrach
(Supervisor)
R. Silberstein
E. Benesch
I. Baum
J. Blum
P. Board
E. Forrest
S. Mitchell
S. Nichols
K. Stingle

HEAD START LANGUAGE COMPARISON

The purpose of this study was to compare the effectiveness of three different preschool language programs. If a significant difference in improvement in children's language could be attributed to a particular approach, the materials producing the most change in the desired direction can be used in current Head Start programs to facilitate instruction as well as provide stimulation and enrichment.

Three treatment conditions were employed:

> UCLA Preschool Language Program: a set of sequenced lessons designed to teach a wide variety of response classes to preschool children.
>
> Behavioral Research Laboratories (BRL) Readiness for Language Arts Program: a set of group-paced instructional materials to develop reading readiness skills.
>
> Control treatment in which children were given non-directive language experiences such as listening to stories and songs on records, but no specific language instruction.

Two classes at each of four Head Start sites from the same Delegate Agency participated in this study. These centers provided eight classes or a total of 120 children, 40 children in each of the three conditions. Each day the children participated in groups of five and receive approximately 15 minutes of instruction.

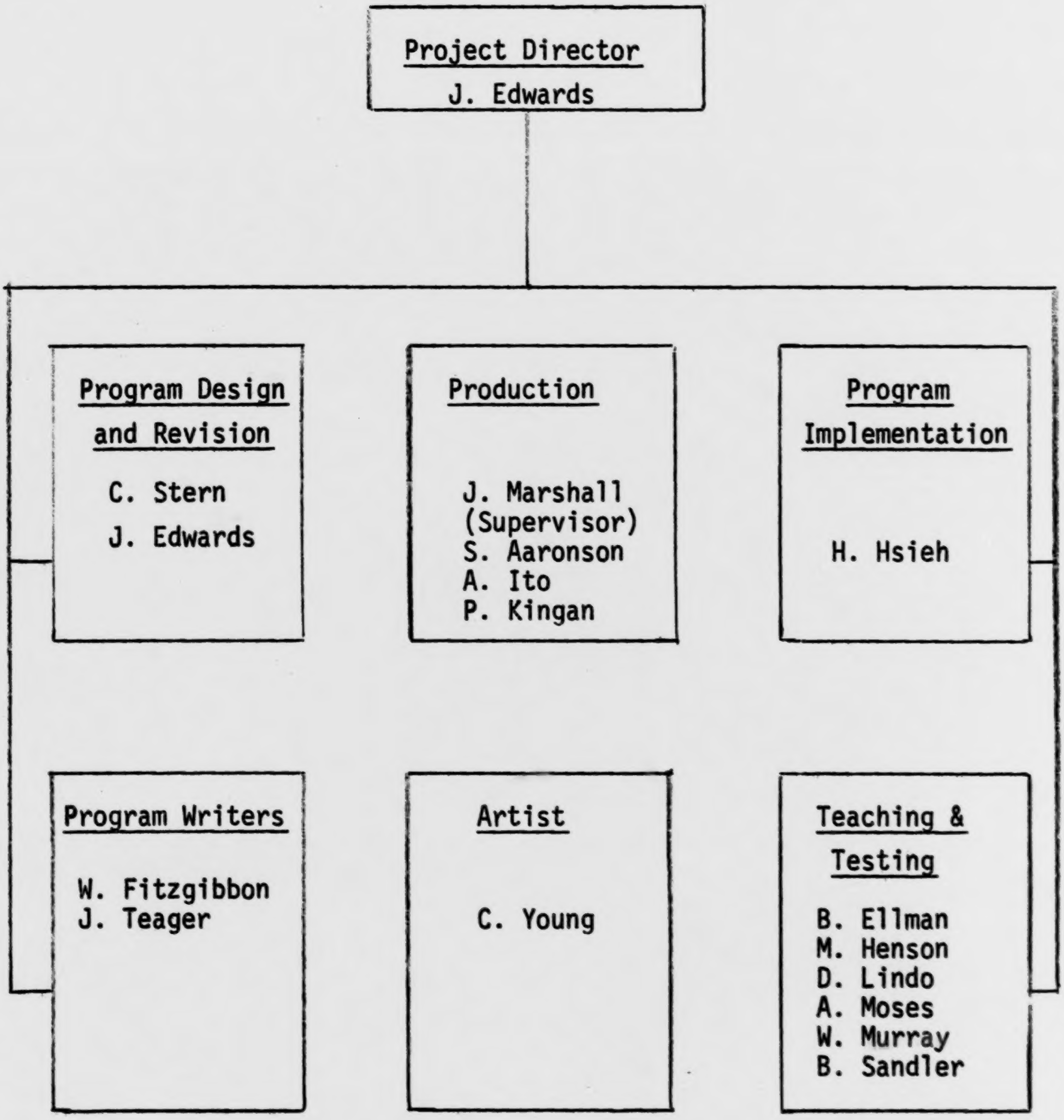
All children were pretested so as to establish baseline performance levels on the behavioral objectives for both instructional programs. The revised Caldwell Preschool Inventory, the Peabody Picture Vocabulary Test, the Gumpgookies Measure of Motivation, and the Visual Discrimination Inventory, were administered as pre and post measures.

The two language programs are sufficiently distinct so that this experiment could be expected to provide guidelines for selection of reading readiness materials for different types of children, so as to add substantially to the overall educational impact of the Head Start Program.

A report of this research project was presented at the Annual Meeting of the National Association for the Education of Young Children in Salt Lake City, Utah in November 1969. A final report has been submitted to the Research Department of the Office of Economic Opportunity.

PS 003298

HEAD START LANGUAGE COMPARISON


Project Director
J. Edwards
Program Design and Revision
C. Stern
J. Edwards
Production
J. Marshall (Supervisor)
S. Aaronson
A. Ito
P. Kingan
Program Implementation
H. Hsieh
Program Writers
W. Fitzgibbon
J. Teager
Artist
C. Young
Teaching & Testing
B. Ellman
M. Henson
D. Lindo
A. Moses
W. Murray
B. Sandler

FOLLOW-UP OF CHILDREN IN THE 1967-1968 EVALUATION SAMPLE

In general, the rationale for the Follow-Up Study was to compare children who had Full-Year Head Start experience under different types of auspices. Thus, in Santa Monica two sites with two classes per site representing Head Start under a Community Action Program, were contrasted with two classes in the San Bernardino (Ontario-Montclair) School District and San Diego (Carlsbad) School District where the Head Start Programs are sponsored by the Local Educational Agency.

In addition, in San Bernardino there was an unusual opportunity to obtain a control from a similar ethnic and socioeconomic background. Here there was a group of children who had been recruited and enrolled in Head Start classes in Cucamonga, but were not given the opportunity to attend because there were severe cuts in the Head Start appropriation and the program was not funded.

Another type of comparison involves Indian children on the Salt River, Arizona reservation. These are children who were in Head Start instead of kindergarten and were in first grade in 1968-1969. A total of 80 children had been tested by the Head Start Evaluation and Research Center at the University of Texas. The Field Supervisor also tried to locate a control group of Indian children in Reservation schools who did not attend Head Start, but very few such children were available for comparison.

PERSONNEL[1]

A. Project Directors

Dr. Joseph Edwards	Language Study	Research Associate
Dr. Barbara Henker	Physiological Correlates of Cognitive Behavior	Professor of Psychology
Dr. Evan Keislar	Research in Cognitive Learning of Young Disadvantaged Children	Professor of Education
Dr. Harry Kitano	Studies in Teacher-Parent Interaction	Professor of Social Welfare
Dr. Millard Madsen	Studies in Cooperation and Competition: Developmental and Cross-Cultural Variables	Professor of Psychology
Dr. Carolyn Stern	Development of Evaluation Instruments	
Dr. Bernard Weiner	Studies in Ascription Behavior as a Function of Socioeconomic Class	Professor of Psychology

B. Follow-Up Supervisors

Dr. Margaret Faust	San Bernardino	Professor of Psychology Claremont
Dr. Sarah Hervey	Salt River	Professor of Psychology University of Arizona
Mrs. Marjorie Royle	San Diego	M.A., Psychology

C. Consultants

Dr. Peter Bentler	Factor Analysis	Professor of Psychology
Mrs. Alita Letwin	Early Childhood Education	M.A., Bank Street College

---

[1] Many of the persons listed were employed for short periods of time and for specific tasks.

D. Center Staff

Dr. Carolyn Stern	Director	
Mrs. Elizabeth Raymer	Evaluation Coordinator	M.A. Candidate, Educational Psychology
Mrs. Allison Burleigh	Administrative Assistant	M.A. Candidate, Educational Psychology
*Mary Avila	Data Aide	Paraprofessional; Nutrition aide in Title I Preschool Program
Susan Aaronson	Clerical and Production	Undergraduate
Arrie Bachrach	OSCI Trainer and Supervisor	M.A., Political Science; two years as observer with UCLA Head Start E & R Center
Steve Baroni	Binet Tester	
Ila Baum	Observer-Interviewer	Paraprofessional; Assistant teacher in Title I Preschool Program
Eva Benesch	Observer-Interviewer	Nursery school teacher
Judi Blum	Observer-Interviewer	Paraprofessional
*Paula Board	Observer-Interviewer	Junior, Sociology
*Shirley Boykin	Data Aide	Paraprofessional; Hospital aide; Volunteer in Head Start class
Harriet Braiker	Research Assistant (Madsen)	Junior, Psychology
Canan Broadbent	Binet Tester	
*Hattie Brown	Data Aide	Paraprofessional; Watts community
Sadelle Brussell	Feedback Analyst	B.A. in Psychology; 1 year special psychiatric preschool internship, Mt. Sinai Hospital

---

*Indicates member of minority ethnic group

*Roland Carter	Binet Tester	
Carol Chambers	Senior Typist-Clerk	
*Sarah Davis	Interviewer	Paraprofessional
Geraldine Devor	Binet Tester	M.A., Educational Psychology
Bruce Ellman	Tester	Senior, Psychology
William Fitzgibbon	Research Assistant	M.A. Candidate, Education
*Elma Forrest	Observer-Interviewer	Senior, Sociology
Aileen Gaal	Research Assistant (Kitano)	M.A., Educational Psychology
Barbara Goetz	Research Assistant (Kitano)	B.A., Psychology
Jeffrey Goldsmith	Coder	Senior, Psychology
Willa Gupta	Data Analyst	M.A., Educational Psychology; Doctoral Candidate in Experimental Design
*Mattie Henson	Teacher, Language Study	Paraprofessional
Diana Homer	Binet Tester	
*Helena Hsieh	Data Records	M.A., Educational Psychology
*Allan Ito	Production Assistant	Junior, Engineering
Margaret Johnston	Senior Typist-Clerk	
Spencer Kagan	Research Assistant (Madsen)	Doctoral Candidate in Psychology
Patricia Kingan	Production Assistant	Junior, Mathematics
Steven Kosslyn	Research Assistant (Henker)	Junior, Psychology
*Robbie Kuykendall	Data Aide	Paraprofessional, Watts Community Skills Program
Carole Langworthy	Binet Tester	
Bernice Ledo	Key Punch Operator	
*Dorothy Lindo	Teacher, Language Study	Paraprofessional
Lynn Lockhart-Mummery	Research Assistant (Weinberg)	Doctoral Candidate, Sociology of Education
David Lourie	Video Cameraman	M.A. Candidate, Theatre Arts
June Marshall	Curriculum materials; Production Supervisor	Nursery School Director

Sandra Mitchell	Observer	Graduate student, Psychology
*Arthurine Moses	Teacher, Language Study	Paraprofessional
*Wanda Murray	Teacher, Language Study	Paraprofessional
Linden Nelson	Research Assistant (Madsen)	Doctoral Candidate, Psychology
Maria Nemeth	Research Assistant (Henker)	Graduate student, Psychology
*Sharon Nichols	Observer-Interviewer	Paraprofessional
Michelle O'Daniel	Data Clerk	Undergraduate, English
Debbie Priddy	Binet Tester	
Linda Reed	Research Assistant (Weiner)	M.A. Candidate, Psychology
Barbara Rosenquist	Research Assistant (Kitano)	Doctoral Candidate, Education
Beatrice Sandler	Teacher, Language study	Paraprofessional
Samuel Schutz	Research Assistant (Keislar)	Doctoral Candidate, Educational Psychology
Ariella Shapira	Research Assistant (Madsen)	Doctoral Candidate, Psychology
Ruth Silberstein	Research Assistant	M.A. Candidate, Educational Psychology
Steven Smith	Coder-Programmer	Senior, Psychology
Donna Stollberg	Coder-Data Clerk	Senior, Fine Arts
Karen Stingle	Observer-Interviewer	Graduate Student, Psychology
Joyce Teager	Research Assistant	M.A. Candidate. Educational Psychology
Dorothy Telson	Senior Typist-Clerk	
Sally Thomas	Research Assistant (Keislar)	Doctoral Candidate, Educational Psychology
*Mireya Valencia	Binet Tester	Doctoral Candidate, Educational Psychology
Michael Vandeman	Programmer	Graduate student, Psychology
John Walker	Research Assistant	Senior, Psychology
*Frank Wang	Data Supervisor	Graduate student, History
*Clarence Young	Artist	

PART II HEAD START EVALUATION AND INTERVENTION PROGRAM

MAXIMIZING THE VALUE OF EVALUATION FOR THE HEAD START TEACHER

Introduction

Many evaluation efforts throughout the country, as well as classroom research experiments in general, report a pervasive conflict between maintaining the on-going program, with the concomitant good will of parents and program personnel, and conducting the required evaluation activities under sufficiently controlled conditions to make the data collected meaningful and comparable. The value of removing children from planned classroom activities for an hour or more at a time so that they may be tested has been seriously questioned. The conflict is aggravated by the fact that the evaluation activities are often unclear, in many cases misunderstood, and almost always non-responsive to the immediate needs of the teacher. With classroom observations, the mystery surrounding the impassive observer with her clipboard and stopwatch creates ambiguities and anxieties which there has been no systematic attempt to allay.

During the development of the classroom observation instrument (OSCI), and in a small pilot study conducted following the 1967-68 evaluation, there was opportunity to converse with Head Start teachers with much more freedom than had been permitted in the sample classes. It was consistently confirmed that teachers, who were disturbed and anxious when they were not informed specifically as to what was going on, relaxed and became more accepting when the objectives of evaluation were clarified. Invariably, they were eager to know how the children in their classes performed on the various tests and when this type of data was provided much of their hostility and suspicion was dispelled.

However, many of the objections and criticisms of specific tests are not without justification; it was found that if the teachers were told that the evaluators were aware of the weaknesses of many of the available instruments, that research teams were exploring new ideas, developing new and, hopefully, better measurement techniques, they tended to become more sympathetic to the evaluation.

Thus, the question of whether time and effort spent in developing a systematic program to provide evaluation information to teachers would pay off in greater teacher understanding and acceptance of evaluation seemed to be worthy of experimental testing. Two basic questions are implicit here: first, would a feedback intervention program designed to make teachers more receptive to evaluation produce measurable changes in teacher attitudes toward evaluation; and second, would these changes in teachers, if they occurred, facilitate the desired changes in child performance?

Accepting the value of a feedback program, the next question concerns the nature of the most effective techniques to use to provide teachers with meaningful information. The value of videotapes in the training and perfection of performance skills has been amply demonstrated in the field of sports. The technique is now being widely applied in the training of teachers. In this area, the evidence (Nimnicht, 1968) indicates that videotapes are effective when they focus on specific behaviors, rather than broad questions such as, for example, classroom management. Providing teachers with the opportunity of seeing themselves on videotape not only serves the purpose of improving teacher behavior, but also provides increased awareness of the value of observational procedures.

Another type of feedback information relates to the performance of individual children on various types of tasks. Datta (1969) carried out a diagnostic feedback experiment, using the regular Head Start teachers employed by the local Delegate Agency. So far, no changes in children attributable to this type of feedback was obtained. In contrast, Spicker, Hodges, and McCandless (1967) found a diagnostic approach effective, but here the teachers were employed by the experimenters who made sure that procedures were modified on the basis of information derived from the diagnostic instruments.

In brief, the UCLA intervention experiment was designed to test whether a combination of feedback procedures, which individually do not seem to have significant effects, would produce differences in teacher behavior and consequent changes in child achievement and socio-emotional growth. The effectiveness of the intervention would be measured first in terms of changes in teachers, and second in terms of changes in children.

Hypotheses

1. Compared to teachers in classrooms where no feedback is given, teachers in the experimental treatment will (a) develop more favorable attitudes towards evaluation as measured by the Teacher Attitudes Toward Evaluation Scale, and (b) be more apt to use a variety of curricular procedures and materials, as measured by range of activities and materials recorded on the OSCI teacher observations.

2. Children in these classes will demonstrate significantly greater gains on both cognitive and affective measures than children in the control classes.

3. Teachers given feedback will have a high degree of correspondence between their ratings of the importance of certain behavioral goals for Head Start classes, the estimates of their achievement of these goals, and the actual performance of children on tasks representative of these goals.

To test these hypotheses, 24 classes were selected from four different Delegate Agencies and assigned randomly, by site, to one of the three treatment groups: Experimental (Feedback), Control 1 (complete test package), and Control 2 (minimal testing, no feedback). The outline of the experimental design and allocation of classes to treatments are presented in Table 1. A more detailed description of each site is included on the following pages (15-21).

GENERAL DESCRIPTION OF DELEGATE AGENCIES AND HEAD START SITES

Delegate Agency I:

Site IA was located in the inner city area. The immediate neighborhood was one of older, single-family dwellings, some newer two and three story apartments, and small old business establishments. There were four classes held in two portables. Each portable consisted of one large room subdivided into two work areas by mutual agreement rather than an actual physical divider. The classrooms were bright and cheerful, well-lighted, and with excellent space utilization. Each portable had its own bathroom facility. The portables were located on a large fenced lot which provided play area front and back. There was a large, grassy front yard, which had a set of parallel bars and some heavy, rubber tires as the only stationary large-muscle equipment. Tables were placed convenient to the sand and water areas. There also was a small row boat, which was filled with water and served as a home for the duck who was the school mascot. The play area at the back of the site was primarily cement, although there was a sand area under the double swing set. A small slide completed that area's large-muscle equipment. These yards were shared by two classes at a time on a rotating basis. Almost all the children were Negro; only three of the 60 enrolled in the four classes were of Mexican-American origin. Of the four original Head Teachers, three transferred to other sites or positions during the year. One class had no assistant teacher; in the other three classes the same assistants stayed for the entire school year. Even with the change in personnel, there was always a three-to-one ratio of Black and Caucasian Head Teachers. As a whole, the staff was well-trained, cooperative, and flexible. A few parents were occasionally involved in the class activities; there were no regular volunteers.

Site IB, also in the inner city area, was located in a predominantly business area. The environment is superficially middle-class, with large homes surrounded by well-kept grounds; however, most of the old mansions have been converted to other purposes and no longer house single families. The two Head Start classes were located in such a converted mansion, on the corner of a busy street. One class used the former living room, the other what might have been a family room or dining room. Although there was a large, sunny kitchen where snacks were prepared, lunches were brought in from another site. The house contained three bathrooms. A former bedroom was used as the office, another as a storage room. The limited space was well-used; separate small areas were set aside for a house-keeping corner, block area, etc. Except for one Mexican-American child, the children were all black. The initial staff was also all Negro. However, one Head Teacher, who had no assistant from the beginning of the term, transferred to another site in mid-term. She was replaced by a Caucasian, with no noticeable change in the program content. A community volunteer assumed the role of assistant teacher and participated daily in the program. She acted as nutritionist, teacher, and aide. Only at the very end of the term was she made the official assistant teacher. In the other class, the assistant left and was not replaced until the end of the term. One Negro Head Teacher was cold and openly hostile, even to the Negro members of the evaluation staff. The Caucasian Head Teacher was ambivalent and unpredictable. There were no volunteers other than the one who later became an assistant teacher. Parents did not participate in the classroom activities but did attend afternoon meetings on issues concerning the welfare of their children.

Delegate Agency II:

Site IIA was located at the very edge of the city, about 20 miles from the downtown area. The physical appearance of the community presents a picture of housing projects, small neighborhood stores, and small and large industrial plants close by. Public transportation into the heart of the city is limited. There is a feeling of transiency and alienation from the larger community. The two classes were housed in the familiar portable unit. The classrooms were bright, colorful and well-planned to take maximum advantage of space and equipment. There were two bathrooms, one at either end of the portable. A small kitchen area provided facilities for warming lunch and/or preparing snacks. There was adequate storage space inside but only limited space outside. There was a small grassy area that was available for easel painting and water play. There was also a tree with pulley attached (used for scientific exploration) and a swing set nearby. Tricycles, a barrel, slide, and sand box were available. The classroom population was totally Negro, with many more boys than girls. Of the two Head Teachers, one was Negro and the other was Caucasian. Both assistant teachers were Negro. The teaching staff was exceptionally cooperative and concerned with a good cognitive program. Principles of positive reinforcement were utilized to maintain active involvement in the learning situations. There was good participation in the parent education program as well as in the classroom. In addition, community people were often brought into the classroom to meet the children. One such visitor was a young artist who brought samples of his work which he discussed. He also talked about his continuing in school to learn more. A very positive community feeling prevailed which manifested itself in contributions of labor and materials to the site and friendly relationships with the staff. Further evidence of the positive, cooperative feeling of the community was the fact that the site never had any thefts or damage to its property. The children registered in these classes lived within walking distance.

Site IIB was located in the southwestern part of the city, a good distance from the inner city area. The building was a modern prefabricated structure, of the type frequently used in California for supplementary classroom space. It consisted of one large room subdivided into two classrooms, which were clean, comfortable, well-lit and with excellent facilities. There was a kitchenette area and bathroom in the back. Lunch was brought in from another site and snacks were prepared on the premises. The building was located in a large park which provided an extra outside space in addition to the fenced-in area immediately surrounding the building. Both areas were used by the two classes, but at different times during the day. There were only a few tricycles and very little in the way of outside equipment, except for a sand box. The park had a large swing set and a merry-go-round. The teaching staff was completely Negro, with one male as a Head Teacher. The staff was competent and warm. There were no volunteers and only occasional parent participation.

Site IIC included two classes in a standard portable unit located on the spacious, well-kept grounds of a large housing development. The surrounding area was primarily one of small businesses, with a few large factories. There was a great deal of interaction between the members of the two classes, who shared both indoor and outdoor facilities in common. The entire teaching staff and the children were Negro; there were approximately an equal number of boys and girls. The atmosphere was warm and child-oriented; the materials were easily accessible and available to the children at their own option.

Delegate Agency III:

One of several autonomous agencies under a large city grantee, this particular agency was associated with a professional early childhood education association whose entire commitment was to this area.

Site IIIA was located on a busy city street. The surrounding neighborhood presented an appearance of modest frame and stucco homes alternating with occasional "papa-and-mama" groceries and small business establishments. All the classes were held in the gymnasium of a large YWCA building. The one large room was divided by a movable wall into two large areas. The space was well-utilized, but inadequate for the number of children housed. A large kitchen was available so that snacks and hot lunches could be prepared on the premises. A single multiple-unit bathroom was at the rear of the building and was shared by staff and children. In this large gymnasium, with its high ceilings and hard-surface floors, there was very little sound-proofing, resulting in an excessively high noise level. Fortunately, right outside was a large playground with a fenced concrete area. Equipment consisted of a swing set, slide, and parallel bars. Trikes and easels were available, and the space was adequate though shared by all five classes. The classroom population was predominantly Negro with a few Mexican-American and Caucasian children. A bus provided for the transportation needs of the site picked up and delivered children, many of whom lived as much as 10 miles away. Only two of the five classes at this site were used in the investigation. One class had a Negro Head Teacher, the other a Caucasian; both assistant teachers were Negro. The Head Teacher of one of the classes was openly hostile throughout the entire study period; the other teacher represented the extreme of cooperation. There was little parent participation nor were there any volunteers.

Site IIIB was located about nine miles south of the center of a large metropolis. The class was held in the anteroom of a church. It was cold, damp, cement-floored, and dimly lighted. Space was woefully inadequate; a small book-shelf, one rectangular table and two small circular ones filled the room. Story-time had to be held in the church proper, which was immediately adjacent. The only source of heat was the gas cooking stove in the kitchen next to the classroom. Food was prepared elsewhere and delivered to the site. The outdoor play area was fenced cement. Play equipment included a make-shift "house", used for dramatic play, and a set of parallel bars. A small house, part of the church and located on the same property, was used for the storage of tricycles and wagons. The outdoor area was really the center of all activities such as water play, easel and fingerpainting, and dramatic play. The classroom population was totally Negro. The children came from the immediate neighborhood, which was similar to that described for Sites IVA and IVB. Both the Head Teacher and her assistant were Negro. The teachers were pleasant, attractive, and responded to the individual needs of the children. Most of the daily indoor time was spent in structured cognitive learning, with greatest emphasis on shapes, colors, and concept labels. There was little parent participation nor were there any volunteers.

Site IIIC was located in the inner city area of a large metropolis. The district was a run-down, low-income residential area. One portable, with no stationary dividers, housed the two classes. These were separate entities only for bookkeeping purposes; in actuality they intermixed freely and groupings were based on activities rather than assignments to a particular teacher. There were special activity areas such as a doll corner for dramatic play, a learning center with books, records, games, etc., and a block-building area. The portable was located on a large, grassy, fenced lot. A dirt play area contained a jungle gym, swing set and slide. A variety of wheeled toys were available. Easels were set up outside so that art activity was available during the free play period. There was a large sand-box and area for water play. At the front of the site was the children's flower and vegetable garden. Rabbits were kept outside. The outdoor facilities offered not only pleasant but also varied and stimulating learning opportunities. The atmosphere was warm and cheerful. There were two bathrooms; the kitchen where snacks and hot lunches were prepared was clean and had adequate space. The class contained an approximately equal number of Negro and Mexican-American children, with a few Caucasian and one American-Indian child. At the beginning of the school year, the staff consisted of two Negro Head Teachers, one Negro and one Spanish-speaking assistant teacher. Because of illness, one of the Head Teachers was replaced about mid-term by a Caucasian teacher.

Although the staff was competent and cooperative with the intervention personnel, there were severe personality conflicts among them. When these problems were brought to the attention of the supervisor she called a meeting and invited airing of grievances. After this session the atmosphere was considerably improved. There was a great deal of personal involvement in the home life of the children, to the extent of providing direct assistance when there was a physical or financial problem. As could be expected, this engendered a very active volunteer program with a majority of the parents involved in school activities. They made a storage shed for wheeled toys, participated and contributed in social gatherings held at the school, formed a cooperative baby-sitting service, and acted in a unified, cohesive way when aid of any kind was needed by a member of the community. For example, the parents provided furnishing for a community member's home, brought clothing for families more needy than themselves, aided in transportation to school, provided child care for families with ailing parents and helped meet the medical costs of a community member's hospital bills. There was a prevailing, solid community feeling centered around the school. The community was extremely involved and protective of the school; for example, one night, a light was inadvertently left on in the building and a community member called the police to survey the situation.

Delegate Agency IV:

This agency was one of four small Community Action Programs in a large Metropolitan area. It was a cooperative venture involving three independent communities.

Site IVA was located in a church building in a city of about 75,000, about 10 miles from the center of a large metropolis. The neighborhood environment looks like a semi-rural small town which has been by-passed by super-highways. Interspersed among the houses, which are small and dilapidated, with peeling paint and slipping shingles, are small factories of crumbling brick or scabrous stucco. The population is predominantly black with a large minority of Mexican-Americans.

There were two classes at this site, both sharing a long, narrow room attached to the rear of the church proper, with not even temporary dividers to separate the classes. One bathroom served both the staff and children of the two classes. A small kitchen, where snacks and lunches were prepared, was directly off the classroom which was well-lit but dirty, crowded, and redolent of human odors and memories of meals long since consumed. To say that the indoor space was inadequate would be a gross understatement.

A large, grassy area was used for outside play activities. The property was fenced off from the adjacent lot, though the driveway leading to the play area was unfenced. Play equipment consisted of a jungle gym and a swing set, both in a sad state of disrepair. The play surface, patches of dirt with occasional swatches of stubborn grass, was hole-pocked and strewn with lumber, broken glass, car parts, and skeletons of wrecked cars. Children moved from indoor to outdoor activities with very loose adult control. Reflecting the neighborhood ethnic mix, most of the children were black, one assistant was black and one was Mexican-American. Three young adults, paid by Work Study funds, occasionally participated in the program. In addition to these paid staff, many parents acted as volunteers on a regular basis.

Site IVB was located in a portable unit in the same small town as Site IA; the neighborhood of this site presented the same shabby combination of rural-industrial structures. Small, run-down single family dwellings were mixed with manufacturing plants and heavy industry. Although two classes shared the same physical area, one was held in the morning and the other in the afternoon, so that the total facilities were available to each class in turn. Morning and afternoon snacks were prepared on the premises, but the hot lunch was brought in from another site.

The indoor environment was large and bright but so fastidiously neat as to create an impression of bareness. There was a doll corner, a block area, and sufficient storage for equipment and books, but the equipment went unused most of the time. There was a high degree of teacher control; the atmosphere of the class was usually very hushed, with the children engaged in quiet free-play activities. Unlike the first site, there were an approximately equal mix of Negro and Mexican-American children, most of whom lived within walking distance.

There was some overlap between the two classes, with the afternoon class arriving at lunch time while the morning class was still in session. Although this was not set up as a full day program, several children remained for both sessions. All four teachers were at the site together most of the day and took turns supervising the classes. The teachers were apathetic and rarely involved on any level other than a disciplinary one with the children. There was one Negro and one Mexican-American Head Teacher. Both assistant teachers were Negro. There were no volunteers, no supplementary paid adults as in Site IV A, nor was there any active or consistent participation by parents. On most pay days of the year, either no classes were held or the children were dismissed early.

The third class at Site IVB, although sharing the same outdoor facilities, was housed in its own portable unit and did not have to share with another class. The classroom was extraordinarily bright, cheerful, and very child-oriented. There was a high tolerance for messy activities, with both painting and water play permitted inside. Separate areas were set aside for dramatic play, science, music, and rest time. The pervading feeling was one of color, warmth, and openness. There were two outdoor areas for large muscle activities. The one in front of the portable was a grassy area (planted by parents of a previous Head Start class) and had one cement pipe through which the children could crawl. The second space was behind the portable and consisted of both a grass and a sand box area. with a large tree which provided welcome shade. There was no other large-muscle play equipment. Only occasionally were the few tricycles taken out of storage and made available to the children. Since the outdoor area was shared by all three classes, it did not provide sufficient space per child. With the exception of one Negro child, the class was made up of Mexican-American children from non-English speaking homes.

At the beginning of the school year, the Head Teacher, a Mexican-American, was promoted and her Negro assistant assumed the responsibilities of Head Teacher, though she was not granted Head Teacher status. This "acting Head Teacher" was totally involved with the children and responsive to their needs in spite of her lack of fluency in their native language. There were no volunteers at this site, and only on rare occasions did a parent participate in the classroom program. The children lived within walking distance and were usually brought to school by mothers or older brothers and sisters.

Site IV C was located in the same small community, but the immediate neighborhood, with its well-maintained single-family dwellings and well-kept gardens, presented a superficial appearance of a quiet, middle-class residential area. Two classes were housed in a large old mansion which had been converted into a church, with accompanying offices and meeting rooms. The building was set back from the street and fronted by an expanse of well-manicured green lawn. One class was located in a cold small room, lit by one bare bulb. Despite the lack of facilities, the room was cheery, with excellent space utilization, adequate table space, and storage facilities. It was immediately evident that this was a room planned for and used by children.

The outside play area, a large, grassy expanse fronting the church, was sometimes used for free-play activities by the children. Near the grass was a dirt area where a large swing set had been set up. There was no other large-muscle play equipment. A cemented area at the back of the enclosed grounds was used for easel painting and water-play activities. Trikes and dress-up materials for dramatic play were regularly made available in this outside area. The classroom population was predominantly Negro with a few Caucasians. Lunch and snacks were prepared on the site. The lunches were varied, tasty, nutritious, and plentiful. The children were bussed in from outlying areas; since one bus served three different sites, the school day often was late in starting. The Head Teacher was Negro and her assistant was Mexican-American. The staff, warm, competent, and directly involved with the children, generated a good learning atmosphere.

The second class at this site shared the same outdoor areas but its indoor facilities were entirely different. The classroom was located in a small room attached to the rear of the church. The room was clean, bright, and cheerful. Books, science equipment, and music equipment were easily accessible and the children could use them at will during their free-play periods. The classroom population was predominantly black, with only a few Mexican-American children. At the beginning of the school year, the Head Teacher was promoted to supervisor and the Negro assistant teacher assumed her responsibilities. Both she and her assistant teacher, also black, were very receptive to the children. However, in spite of the excellent teaching environment, there were no volunteers and very little parent involvement.

Table 1

EXPERIMENTAL DESIGN AND ALLOCATION OF CLASSES TO TREATMENTS

	Treatments			
Design Components	Experimental (A)	Control 1 (B)	Control 2 (C)	Total
Agency I	4	2	0	6
II	2	2	2	6
III	2	1	2	5
IV	2	3	2	7
Total Number of Classes:	10	8	6	24
Total Number of Teachers:	20	16	12	48
Intervention	Feedback Meetings; Videotaping	Orientation and Final Meetings Only	None	
Test Battery	Total evaluation package of common core instruments, individual child tests, classroom observations and interviews with parents and teachers, as prescribed for the national evaluation, plus instruments designed especially to test the hypotheses of this study.		Minimal pre-post testing of sample children (PPVT, not Binet). No observations, and no parent interviews.	

Subjects

Selection of Sample Agencies, Classes, and Children.

Since the evaluation program for the proposed intervention was to involve a large number of classes and a great deal of close interaction during the year, it was decided to limit the geographic area from which the sample was selected. Two Community Action Programs with Head Start classes within Los Angeles County were invited to participate in the evaluation program. The largest CAP, the Economic and Youth Opportunities Agency of Los Angeles (EYOA), includes 12 different Delegate Agencies with approximately 6600 children in full year Head Start programs. The UCLA Head Start Evaluation and Research Center Director and the Evaluation Coordinator presented the intervention plan at a meeting with the Director and Assistant Director of EYOA and the Head Start Project Directors and Child Development Supervisors of the 12 Delegate Agencies. Page 24 presents the handout prepared for this meeting.

Many more agencies expressed a desire to participate than could be included in the study. Thus the selection of the three Delegate Agencies for participation in the intervention program was based on the interest expressed by their Project Directors and the order in which they volunteered for the program.

In addition to the large EYOA grantee, a newly-formed small CAP agency was invited to take part. In each case, the Head Start Director for the particular agency selected the sample classes. However, the random assignment of classes to treatments was made by the UCLA staff.

In all of the 24 classes selected at least two-thirds of the children were "sample eligible" as defined by the OEO criteria. These require that (a) the child have no previous Head Start experience, (b) be of pre-school age, and (c) speak and understand English sufficiently well to be tested in English. At one of the sites included in the sample, one of the two classes listed seven eligible children and the other 13. These two classes operated as one large class, so the ratio was computed as 20 sample eligible children of the total 30 enrolled, thus fulfilling the required 2:3 proportion.

Tables 2a and 2b describe the race, age, and sex of the sample population by classes, agencies, and treatments; Table 3 provides the attendance and attrition rates for the same grouping.

SUMMARY OF PROPOSED HEAD START EVALUATION PROGRAM, 1968-69

The UCLA Head Start Evaluation for the coming year has the promise of being both exciting and innovative. In addition to the "common-core" measures required for the National Evaluation, this center will initiate an experimental program of feeding back information about the results of the program directly to teachers in approximately half of the sample classes. By making information available to teachers during their programs rather than months after these programs have ended, it is hoped that:

1. Teachers in "feedback" classes will develop a more favorable attitude to evaluation than teachers in classes where no feedback is given.
2. Teachers receiving feedback from classroom observations and video-tape sessions of their classes will be more apt to use a wider variety of curricular procedures and materials than those who do not receive feedback.
3. Children in the "feedback" classes will demonstrate significantly greater gains on both cognitive and affective measures than children in the control classes.

It should be noted that this is not a training program and will in no way interfere with the plans of individual agencies for teacher training. Classes involved in the evaluation program should not, however, be a part of any other evaluation or research program. Project Directors and Child Development Supervisors of Delegate Agencies participating in the Evaluation Program are invited to attend any teacher feedback meetings and will be kept informed of the progress of the Program.

Table 2a

Race, Sex, and Age of Sample Population, by Class

Agency	Treatment	Class No.	N	Race				Sex		Age in Months as of October 1, 1968				
				W	N	M-A	Mixed & Other	M	F	45-7	48-50	51-3	54-6	57-9
I	A	1	10	-	10	-	-	5	5	-	8	1	1	-
I	A	2	9	-	8	-	1	8	1	1	2	2	4	-
I	A	3	14	-	14	-	-	4	10	1	6	1	3	3
I	A	4	13	-	12	1	-	7	6	1	3	2	6	1
II	A	5	10	-	10	-	-	5	5	1	2	3	3	1
II	A	6	10	-	10	-	-	8	2	3	3	3	1	-
III	A	7	15	-	10	4	1	7	8	5	3	4	3	-
III	A	8	15	-	10	4	1	9	6	1	5	5	3	1
IV	A	9	12	1	11	-	-	8	4	3	2	3	3	1
IV	A	10	13	1	10	2	-	8	5	3	4	3	2	1
I	B	11	15	-	15	-	-	6	9	1	3	7	4	-
I	B	12	14	-	14	-	-	7	7	2	5	4	2	1
II	B	13	13	-	12	-	1	7	6	4	4	2	2	1
II	B	14	7	-	7	-	-	4	3	2	2	1	1	1
III	B	15	13	-	13	-	-	7	6	1	3	5	3	1
IV	B	16	11	-	3	8	-	6	5	-	2	2	6	1
IV	B	17	14	-	6	8	-	10	4	1	3	2	7	1
IV	B	18	13	-	6	6	1	3	10	1	2	3	6	1
II	C	19	12	-	12	-	-	7	5	2	5	3	-	2
II	C	20	10	-	10	-	-	4	6	1	4	2	2	1
III	C	21	16	3	5	7	1	10	6	-	9	5	1	1
III	C	22	13	1	5	6	1	8	5	3	4	6	-	-
IV	C	23	10	-	9	1	-	4	6	4	1	2	2	1
IV	C	24	14	-	14	-	-	8	6	3	3	1	3	4

Table 2b

Race, Sex, and Age of Sample Population, by Treatment

(With Attrition based on Pre-Post Binet or PPVT)

		Treatments						Total	
		Experimental (A)		Control-1 (B)		Control-2 (C)			
Variables		N	%	N	%	N	%	N	%
Race	White	2	1.7	0	0	4	5.3	6	2.0
	Negro	105	86.8	76	76.0	55	73.3	236	79.7
	M-A	11	9.1	22	22.0	14	18.7	47	15.9
	Mixed	2	1.6	0	0	1	1.3	3	1.0
	Other	1	0.8	2	2.0	1	1.3	4	1.4
	Total	121		100		75		296	
Sex	Male	69	57.0	50	50.0	41	54.7	160	54.0
	Female	52	42.9	50	50.0	34	45.3	136	45.9
Age in months as of 10/1/68									
	45-47	19	15.7	12	12.0	13	17.3	44	14.9
	48-50	38	31.1	24	24.0	26	34.7	88	29.7
	51-53	27	22.1	26	26.0	19	25.3	72	24.3
	54-56	29	24.6	31	31.0	8	10.7	68	23.0
	57-59	8	6.6	7	7.0	9	12.0	24	8.1
Binet	Pre N	103	84.4	79	79.0	64*	85.3	246	83.1
	Post N	75	61.5	57	57.0	51*	68.0	183	61.8
	% Attrition	28	27.2	22	22.0	13*	20.3	63	21.3

*Peabody Picture Vocabulary Test

Table 3

Percent Attendance and Attrition by Class, Agency, and Treatment [a]

Class							Agency & Treatment					
Agency	Treatment	Class	Percent Attendance		Percent Attrition	No. of Months	Agency	Treatment	Percent Attendance		Percent Attrition	No. of Months
			M	SD					M	SD		
I	A	1	88.5	5.6	30.0	9	I	A	80.3	11.4	36.2	7.5
I	A	2	84.3	8.0	50.0	7	II	A	88.7	6.3	10.0	9.0
I	A	3	65.3	14.6	35.0	5	III	A	82.8	10.3	33.3	9.0
I	A	4	77.4	5.0	26.0	9	IV	A	85.7	9.4	32.0	9.0
II	A	5	89.9	6.5	10.0	9						
II	A	6	87.5	5.7	10.0	9	I	B	78.3	9.7	27.6	8.5
III	A	7	74.9	3.8	13.3	9	II	B	89.8	8.6	35.0	9.0
III	A	8	90.6	8.8	53.3	9	III	B	85.6	6.2	38.5	9.0
IV	A	9	90.5	4.0	16.7	8	IV	B	84.3	7.9	35.1	8.7
IV	A	10	81.8	10.6	46.2	10						
							II	C	No Data			
I	B	11	81.5	6.0	20.0	9	III	C	80.5	10.5	34.6	10.0
I	B	12	74.6	11.9	35.7	8	IV	C	84.3	9.7	8.3	9.5
II	B	13	86.7	6.6	30.8	9						
II	B	14	93.0	9.1	42.9	9	I		79.6	10.9	32.9	7.8
III	B	15	85.6	6.2	38.5	9	II		89.3	7.5	22.5	9.0
IV	B	16	88.9	2.8	45.5	6	III		82.4	9.9	34.8	9.4
IV	B	17	84.6	8.6	42.9	10	IV		84.7	8.9	25.6	9.0
IV	B	18	81.3	7.9	16.7	10						
								A	83.8	10.3	29.5	8.4
III	C	21	81.4	10.4	42.9	10						
III	C	22	79.1	10.5	25.0	10		B	84.4	9.3	33.3	8.8
IV	C	23	91.8	4.7	8.3	10						
IV	C	24	76.2	6.5	8.3	9		C	82.4	10.3	22.0	9.8

[a] The "No. of Months" column refers to number of months for which attendance data were available. There were no records for the C treatment (classes 19 and 20) in Agency II. Totals for Agencies III and IV include C treatment. Agencies II, III, and IV had a 10-month program; Agency I closed their Head Start classes one month earlier.

Criterion Measures

All children and classes in the national evaluation sample were assessed using a common battery of tests and observation procedures adopted for the national evaluation by the Center Directors after a number of planning conferences. The data obtained described variables within three broad categories: individual children, educational environments, and family relationships. To provide consistent and comparable data for the total national evaluation, the core instruments were administered following common guidelines and, in some cases, with all testers trained by the same trainers. Most of the instruments were given as pre-post measures, with several of the observational assessments repeated a number of times during the course of the academic year.

In addition, each of the national Head Start Evaluation and Research Centers was responsible for collecting data with the forms developed under the Bureau of the Census, the general data collection agency for OEO. The Evaluation and Research Center Directors were given this task in order to determine with some reliability the extent to which the classes included in the national evaluation resembled the 5% sample of all Head Start classes randomly selected by the Bureau of the Census.

The Census forms included:

CAP-HS Form 35S - Class Register

CAP-HS Form 31 - Medical/Dental Information

CAP-HS Form 43 - Staff Member Information

CAP-HS Forms 45 and 45a - Parent Participation Record. (The Parent Participation Record consisted of two forms. The Center Activities Record was filled out by the parent activities coordinator; the Classroom Record, which was expected to reflect parent participation in class activities, was filled out by the teacher for each classroom.)

CAP-HS Form 52 - Center Facilities and Resources Inventory

CAP-HS Form 46 - Family Information Form

In many cases, the Census data were similar to and overlapped the information obtained with the instruments used in the national evaluation. As far as possible, the forms were filled out by the person designated, with the assistance of the data aide assigned to the site. When necessary, the Evaluation Coordinator assisted in obtaining the data, e.g. the Medical/Dental Information.

The common core measures administered to the national evaluation sample are described below under the appropriate categories.

I. Individual Child Measures

A. Cognitive Domain

1. *Stanford-Binet*. In spite of a great deal of criticism, this standard measure of intellectual functioning was used as the basic assessment of cognitive ability. The Stanford-Binet is a well-known test and needs no special description here. It should be noted, however, that contrary to the usual procedure, all parts of all items between basal and ceiling were given. In the UCLA Feedback Intervention, for purposes of interpretation to teachers, several subtest scores were derived. The items included in these subtests are listed in detail on page 57. Because the second control treatment received the Peabody Picture Vocabulary Test instead of the Binet, the vocabulary scores on the Binet were also computed separately so that comparisons between the groups could be made with somewhat greater confidence. The table reporting the data from the Binet provides subtest scores as well as gains between pre- and posttest.

Two additional assessments were made concomitant with the administration of the Binet. The first was an elaboration of the original face sheet of the Binet protocol. It was developed to record some of the affective behaviors of the child during the test-taking period. This instrument, called *Factors Affecting Test Performance (FATP)*, provides a rating scale for reporting behavior of the child with relation to the test items as well as to the examiner; it also includes items relevant to the generalized activity and verbal level and a description of test conditions and language environment.

The second assessment which accompanied the Binet was an adaptation of the *Birch-Hertzig Response Style Scoring Procedure*. It was used to record whether or not the child attempted to respond to each item, whether his response was verbal or nonverbal, whether his response was specifically limited to the task or whether he offered spontaneous elaborations. In the "non-work" category, the coding indicated the methods used by the child to avoid responding. These could be substitutions, requests for aid, negation, or passivity. The number of prompts or cues given by the examiner to get the child to respond were also noted.

Since no norms were available for interpreting scores on either the FATP or the Birch, the data from these instruments were not used in the analysis section of this report. However, the performance of the individual children on each of the items has been keypunched and submitted to Washington for the national evaluation analyses.

2. *Caldwell Preschool Inventory (PSI)*. This instrument was originally developed in response to the need to measure the

outcomes of the summer 1965 Head Start program. Since then it has gone through several revisions, mostly in terms of decreasing the number of items and rearranging their order of presentation. The version used in this evaluation was developed as a result of factor analyses based on the data collected for the 1966-1967 national Head Start evaluation. The items on the PSI were also grouped into subtests (see page 58) for feedback to the teachers in the experimental treatment.

3. Animal House (subtest of the Wechsler Preschool Primary Scale of Intelligence) was recommended for inclusion in the battery as a result of a special study to find measures which would be more appropriate for the Head Start population. The Animal House subtest is a categorization task. In each square of the matrix board there is a hole and a picture of one of four animals. The child must associate a color with each animal and place a small cylinder of the proper color into the hole under each animal. This task requires the child to associate sign with symbol and is used as a measure of learning ability. According to the manual, "Memory is, of course, a basic factor, but attention span, goal awareness, and ability to concentrate may also be involved." Scores are based on time required as well as number of errors and omissions.

4. TEACH-Tasks. In addition to these common core measures, a new instrument was developed to assess the degree to which the teachers' estimates of children's achievement corresponded with actual performance on specific tasks. A set of items was devised to sample the behavioral objectives included on the TEACH (see description of this instrument below). A detailed report on the pilot test of this measure is included in the Appendix.

B. Affective Domain

1. Factors Affecting Test Performance (FATP). (See above.)

2. Birch-Hertzig Response Style Scoring Procedure. (See above.)

3. Gumpgookies: A Test of Motivation to Achieve, was developed at the University of Hawaii by Bonnie Ballif and Dorothy Adkins. The Gumpgookie is a cartoon figure that looks something like Casper the Ghost. The child is told that he has his own little Gumpgookie who does what he does and likes what he likes. Each item consists of a pair of these figures. The child is told a story about the item, in which one Gumpgookie behaves in an approved or motivated manner and the other Gumpgookie in a disapproved or non-school-motivated manner. The task for the child is to select the Gumpgookie who is behaving most like the way he would behave in that situation.

This test had had considerable pilot-testing in Hawaii, and had demonstrated an acceptable level of reliability as well as concurrent validity. However, almost all the testers using the instrument with Mainland Head Start children found many problems in its administration. Primarily, the instrument, which consisted of 100 items, was much too long. As a result, children became extremely restless and refused to attend to the pictures before making a choice. Many protocols also showed position-set response patterns. Another common criticism was that the language used in the stories was not relevant to the experiences of this population. To minimize some of these problems, the evaluation coordinators and their testing staffs were requested to rate the various items in order of acceptability for retention in a shortened form. Based on these ratings, 55 items were identified for use in the posttest. The pretests were rescored so that only the 55 items used on the posttest were included in the final analyses.

4. Sociometric: Play Situation Picture Board was developed at Michigan State University under the supervision of Robert P. Boger. This instrument requires that a Polaroid picture of each child in the class be taken just prior to administering the test. These pictures are to be mounted on white fiberboard in four rows of four photos. During the test, the child is asked to find his own picture and then to name the other children. If he cannot name the children, the examiner asks the child to point to the picture as the name is given. After the child has demonstrated that he knows the other children in the class, he is shown a sequence of five pictures involving diadic play situations and is asked to select three situations. When the child has selected three of the five play situations he is asked to select a picture of a child he would like to have engage in that activity with him. Parenthetically, it should be noted that there are actually six pictures, but the doll picture is shown only to girls and the ball picture only to boys. This feature was upsetting to a number of child development supervisors who do not approve of sex-typing of play materials.

There were several serious problems inherent in the use of this instrument. In the first place, obtaining the Polaroid pictures of all the members of the class within a limited period (let alone on the same day as prescribed by the manual) was a virtual impossibility. After the photographer had returned to classes looking for missing children over a period of about four weeks (a most costly procedure) it was decided to go ahead with the testing even though several children had not been photographed.

A second serious problem was that the test provided that only children in the same class be included on the picture board. However, many of the centers in the sample had two or more classes, with one site having five classes all in one large room. In some cases all the children were in the sample, but more frequently they were not. Thus, presumably, a child in the sample not chosen by anyone else in the sample might have been chosen frequently if non-sample children were given the sociometric test. Also, children who were chosen might not have been selected if other more-favored alternatives among non-sample children had been included on the picture board.

The use of the instrument as a mid- and post measure is also questionable. As a change indicator, it is obviously inadequate, since the child who is most frequently selected cannot improve his position and can only lose status; the relative position of each child is highly dependent upon a stable classroom population, which was certainly not the case in the Head Start classes in this sample. Finally, the reliability of the measure has never been established.

Because of the various types of difficulties described above, the data from this instrument was not used in the final analyses. The results were keypunched and sent to Washington for use in the national evaluation.

C. Medical. UCLA was one of the Evaluation and Research Centers who expressed willingness to obtain pre-post measures of height, weight, and hemocritic blood analyses. The actual testing was done by the regular medical personnel with whom the various Delegate Agencies already had contractual agreements.

II. Educational Environments

A. Classroom Variables

1. Observation of Substantive Curricular Interactions (OSCI) was developed at the University of California, Los Angeles, under the direction of Carolyn Stern. The UCLA Center was charged with instrument development, training of observers, and analysis of data; and detailed reports on all of these phases of activity are being submitted separately. Briefly, the OSCI is a time-sample procedure which reports, in a codified fashion, the types and frequency of cognitive inputs and social interactions, the structure of the classroom learning environment including size of groupings, the locus of control and level of teacher involvement, and the variety of materials and the extent to which they are used by children.

Four observations were made at specified intervals over the school year; each consisted of four periods of 10 three-minute units or a total of two hours. The data from these observations are being submitted to a complex computer analysis. To facilitate writing the evaluation reports required of the various Evaluation and Research Centers, a procedure for analyzing the data by hand was devised. This procedure is reported in full in the Appendix, and the results of this analysis are used in interpreting the findings of the present feedback intervention.

2. Observer's Report on the Class (ORC). This semi-objective, semi-subjective overall estimate of the intellectual climate of the classroom was developed at the Washington office and used as a post measure only. It consists of a five-point rating scale applied to seven developmental goals: sensory-motor, language, group and social skills, conceptualization, academic skill, self-esteem and motivation. In addition, the observers were asked to indicate whether other goals than the ones listed were emphasized, and whether or not there was any serious type of disruption or reorganization which might have affected the operation of a particular class. Finally, the observers were asked to write a brief evaluative description, in a somewhat anecdotal manner. The data from the seven rated items, as well as some of the more subjective reports, are included in the analysis section of this report.

3. Class Facilities and Resources Inventory. This instrument was an adaptation, developed at the UCLA Center, of the Bureau of the Census Form 52. As modified, it is both a quantitative and qualitative assessment of the physical space and facilities available to children in a particular class and site. Part I of the Inventory consists of a checklist describing the size of the town, under what auspices the program is operated, the general category describing the building in which the class is held (e.g. church, store, school, etc.), how the child gets to school, the kind of spaces used by the class (e.g. bathrooms, gymnasiums, closets, kitchen facilities, etc.); and the nature of the outdoor play area (location, sun and shade, trees, surface, etc.). A separate section is concerned with the quantity and availability of books both in the classroom, in libraries, and to take home. A comprehensive list of 27 types of materials and equipment, including such items as swings, blocks, dolls, puzzles, pets, audio-visual equipment, etc., is provided and the observer is asked to rate each item on two separate scales: first in terms of quantity and availability and second in terms of the condition in which the items are found.

Part II of the Inventory looks at the accessibility and general appearance of the classroom. In the latter category are such items as color of paint, presence of acoustical tile, type of floor covering, ventilation, windows, screens, lighting, heating and cooling provisions, etc. Storage facilities, tables, seating arrangements, and presence or absence of special activity areas such as doll corner, listening center, art area, etc., are also described in this section.

This instrument should provide a wealth of information to describe the wide variation of learning environments which all come under the Head Start umbrella throughout the country. However, for the purposes of this report, only certain items were chosen and ratings assigned to the descriptions provided in the Inventory. The primary criterion for item selection was to enable differentiation among classrooms on variables that had possible relevancy to teacher and child variables, as measured by other instruments. The two secondary criteria for item selection were first whether the information was complete for all the sample classes, and second whether the obtained data were amenable to a rating system. The detailed description and the rating scheme for the ten variables derived in this manner are presented in detail immediately preceding the interpretation of the findings for this instrument.

B. Teacher Variables

1. Characteristics of Teaching Staff is a UCLA modification of the Bureau of the Census Form 43. It attempts to present a description of the background and experience of the teaching personnel in terms of 18 basic items. The first seven items concern such demographic features as sex, age, ethnicity, language, and socioeconomic status. Items 8 through 12 are concerned with general education, training, and academic background. Items 13 through 18 report the amount of work experience with children both prior to and including the present Head Start teaching assignment.

For purposes of the analysis of the feedback study, all the items mentioned above were included. However, several items were combined in the education and experience categories.

2. Checklist of Administrative Variables, developed by Shuell Jones at Tulane University, was designed to record the "administrative and community variables which directly affect the range of teacher responsibility." From a list of personnel, the teacher is requested to indicate who was the person responsible for a variety of decisions and activities, including selection and enrollment of the children; selection of instructional materials and procedures; obtaining parent involvement, social and medical services, etc. The data from this instrument were keypunched and sent to Washington but were not analyzed for this report.

3. Post Program Interview with Teacher was developed as a common measure to assess the achievement of the objectives of the various intervention models across all the classes in the evaluation sample. The rationale for this instrument was to enable a wider comparison of the effectiveness of specific treatments, and to minimize possible diffusion effects between experimental and control treatments in the same Evaluation and Research Center sample. For this instrument, all the Centers were requested to submit behavioral or operational definitions of their intervention programs and these were then coordinated into the final instrument by the central staff in Washington.

4. Post Observation Teacher Rating Scales (POT) were compiled at Syracuse University under the supervision of John Dopyera. It is a compilation of items submitted by several Evaluation and Research Centers to express relevant features of specific intervention treatments. In general this instrument was designed to get at those intangible aspects which characterize the classroom ambience and which the behaviorally-oriented OSCI does not attempt to assess. Several of the most meaningful items of the University of Texas Observer Rating Scale, as determined by analysis of the 1966-1967 data, reworded to provide a frequency of occurrence scale, were also included. These items were concerned with such questions as reinforcement procedures, instructional aids, and teacher language. The POT was completed by the observer after the first visit to the classroom, and repeated after each of the four OSCI observations.

A training session to familiarize observers with the types of events to be rated revealed many inadequacies and ambiguities in the scaling and interpretation of the items. Particularly confusing was the fact that so many times an explanation seemed completely non-responsive to the item it was supposed to clarify and often created more confusion than if no explanation had been given. Another major source of difficulty was the inclusion of a zero code for "No basis for judgment - occasion did not occur." This code permitted the observer to avoid making a judgment in situations where there was ample opportunity for the behavior under consideration to occur. Further confusion was due to inconsistency in the rating scale - in some cases the scale was based on frequency, in others on degree or intensity of the behavior observed. The valued behavior was sometimes indicated by a high rating, sometimes by a low rating, and sometimes by a middle rating.

In their own critique of the instrument, the Syracuse University staff have pointed out that the subjective aspect of POT is the main difficulty in implementing its usefulness. The problem inherent in obtaining inter-rater reliability from two observers, each of whom may have been observing quite different concurrent events, is also noted. Further elaboration was offered in the attempt to clarify some of the most ambiguous items (4, 6, 10, 14, 19, 26, 30, and 31).

To determine the reliability of the items and to devise a meaningful factor structure, Dr. Renato Espinosa of the University of Hawaii Evaluation and Research Center performed an analysis of the POT data from the 12 national centers. It was hoped that such an analysis would generate scores which could then be used in correlational analyses relating teacher behavior with changes in children. After inspection of the raw data, it was noted that more than half the observers gave zero ratings to items 6 and 33. Also, the UCLA Center data contained a significantly greater number of zero ratings than the other centers.

For this reason, the two items and the UCLA data were excluded from the Hawaii analyses and all zero ratings were changed to the mid-point value. The subsequent analysis produced four rather distinct factor loadings. Two of these factors were described as "quality of cognitive input" and "concern for individual emotional comfort," but the other two factors were difficult to interpret. Two subscale scores were computed, averaging over the total number of observations.

Following the suggestions made for using the POT, the Los Angeles data were analyzed at UCLA, first changing all zeroes to the mid-point score (3). Means for each of the five observation periods and a rating for the direction of change, based on the linearity of the slope of the five scores, was developed.

In addition to these common core instruments used by all the evaluation centers, two measures to reflect changes relevant to the UCLA Feedback Intervention were developed and used. These are the Teachers Attitude Towards Evaluation (TATE) and Teachers Expectations for Achievement of Children in Head Start (TEACH). Since the titles of these measures give a good indication of the kind of information which is of concern, no further elaboration will be given here. Detailed information on these instruments is included with this report.

III. Family Relationships

Aside from the forms which were part of the Census Bureau data collection, the Parent Interview was the only instrument used to obtain data on important factors having to do with the physical aspects of the home environment (e.g., type of domicile, number of rooms per inhabitant, condition of repair, facilities available, etc.), the psychological structure (parental relations, siblings, related and non-related others, etc.), and the economic resources. However, this single instrument was in actuality a compendium of items which permitted a comprehensive examination of this complex variable. The interview included items relating to the following parental characteristics: 1) attitudes toward Head Start; 2) participation in the current Head Start child's class; 3) educational opportunities and aspirations for the child; 4) child rearing practices; 5) mother-child interactions; 6) family information and socio-cultural practices; 7) knowledge about the community; 8) involvement in the community; 9) parent opinions; and 10) interviewer's observations. At the meeting of the Center Directors and Evaluation Coordinators after the completion of the pretesting cycle, this last section was severely criticized. It was felt that the observer was expected to make value judgments which were demeaning and destructive of the tenuous rapport which was being so carefully nurtured. As a result, it was not included on the post-test form of the interview.

The Parent Interview was intended as a pre-post measure. However, the difficulty involved in preparing the final version of this instrument, which included the addition of items to reflect features of specific Evaluation and Research Center Intervention studies where degree of parent involvement was a treatment variable, caused a postponement in the original testing schedule. Another reason for delay was the requirement that the instrument be administered only after the examiners had had an intensive training course by representatives from the Washington office. At UCLA this workshop-type training took less than the two days allocated, and there was confidence that even the paraprofessionals with little academic training would be able to collect reliable data. However, staff training continued even after the Washington trainers continued on their tour of the Centers. One of the trainers returned to provide a refresher session just before the posttest period.

The Parent Interview took about an hour to administer, and was carried out in the child's home. Since no norms were available for this instrument, it was not used in the present report.

Procedure

The entire focus of the intervention treatment was centered on developing in teachers of sample classes a more favorable attitude toward classroom observation and student testing. There was specifically no intention to infringe on the teacher-training function of the child development supervisor, and every effort was made to avoid slipping over into a didactic role. If any teaching skills were learned as a result of the feedback activities, it was due entirely to the inferences and applications of the teachers themselves and not from the UCLA Evaluation staff. The intervention consisted entirely in meeting with the teachers, telling them the type of activities, instruments, and observations that would be part of the evaluation process, and what kind of information the various types of measures were expected to produce. After the testing, additional meetings explained in very general terms the kinds of data that had been collected on the individual children, and also provided teachers with the opportunity to view themselves as they interacted with the children during videotaping sessions.

The following section provides a detailed record of the intervention program. Page 38 contains the Schedule of Activities for the nine months of the study; page 39 shows in tabular form the testing and feedback schedule. After a brief description of the general procedure and the orientation meetings on pages 40 and 41, the next section (pages 41 to 54) is devoted to summary reports of the feedback meetings.

During the course of the feedback meetings, a number of handouts were distributed to clarify the various instruments and procedures being discussed. These included an overview of the total evaluation package (page 55), a brief discussion of the classroom observation (page 56), a breakdown of subtests for the Binet (page 57), and for the Preschool Inventory (page 58). An integral part of the intervention was the use of data aides. A discussion of this program is provided on pages 59 to 62.

SCHEDULE OF ACTIVITIES
1968-1969

September: Select 24 sample classes with 10 eligible children per class. (To be eligible, children must be 3-6 to 4-6 and in the program new.)

Orientation Meeting #1
- Distribute Parent Permission slips;
- Meet with Head Teachers to explain program;
- Administer following questionnaires to teachers:
- TATE (Teacher Attitudes Toward Evaluation) and
- TEACH (Teacher Expectations for Achievement of Children in Head Start).

October: Pretest individual sample children during 4th through 7th week of classes.

Individual Child Observation #1
Orientation Meeting #2 (Experimental classes only)
- Discuss classroom observation procedures;
- Parent Interviews (Pre)

November: Classroom Observation #1; Playboard Sociometric
Feedback Meeting #1 (Experimental classes only)
- Discussion of results of pretesting and class observations.

Video-taping #1 (Experimental classes only)

December: Feedback meeting #2 (Experimental classes only)
- Teachers view and discuss video tapes.

January: Classroom Observation #2
Feedback Meeting #3 (Experimental classes only)
- Report on results of classroom observations.

February: Video taping #2 (Experimental classes only)
Feedback Meeting #3 (Experimental classes only)
- Teachers view and discuss video tapes.

Individual Child Observation #2

March: Classroom Observation #3

April: Feedback Meeting #5 (Experimental classes only)
- Teachers view and discuss video tapes.
- Discussion of data from classroom observations.

May: Posttesting of individual children
Classroom Observation #4; Parent Interviews (Post)
Feedback Meeting #6 (All classes)
- Administer TEACH and TATE as posttests;
- Discuss results of individual tests.
- Discuss teacher's reaction to the evaluation program.

TESTING AND FEEDBACK SCHEDULE FOR INTERVENTION PROGRAM

September	October	November	December	January
Orientation Meetings	[a]Stanford-Binet Birch; FATP	PSI Gumpgookies Animal House	Sociometric	
	[b]Individual OSCI #1 P.O.T.	[b]Group OSCI #1 P.O.T. [c]Videotape #1		[b]Group OSCI #2 P.O.T.
	Feedback Meeting #1		Feedback Meeting #2	
	Class Facilities Staff Characteristics	Parent Interviews		Feedback Meeting #3

February	March	April	May	June
Individual OSCI #2 P.O.T.	Group OSCI #3 P.O.T.		Group OSCI #4 P.O.T.	
Videotape #2		Stanford-Binet Birch; FATP	PSI Gumpgookies Animal House	Sociometric
Feedback Meeting #4	Feedback Meeting #5	Feedback Meeting #6	Final Meetings Class Facilities Teacher Characteristics Administrative Checklist	

[a] Experimental and Control 1 received Stanford-Binet; Control 2 received Peabody Picture Vocabulary Test

[b] Observations for Experimental and Control 1 only

[c] Videotaping and Feedback Meetings 1-6 for Experimental only; Orientation and Final Meetings for all groups

MEETINGS AND FEEDBACK SESSIONS WITH HEAD START TEACHERS

I. General Procedure

An initial orientation session to present the general evaluation plan was held at each Head Start site included in the sample. Each teacher was informed as to whether her class belonged to the feedback or control group. Advance notices of the feedback meetings were sent to the teachers, supervisors, and project directors involved in the feedback program. This notice was followed up by a personal telephone call on the day before the meeting. Feedback meetings were open to Head Teachers, assistant teachers, and other agency personnel such as Child Development Supervisors and trainers. There were usually two Feedback Analysts and a Data Aide in attendance from the UCLA Evaluation and Research Center. All meetings took place after Head Start classes had been dismissed. The meetings lasted an average of three hours to allow presentation of information and discussion. A tape recording was made of each meeting.

II. Orientation Meetings

After preliminary introductions, the teachers and assistants were asked to complete the Teachers' Attitudes Toward Evaluation (TATE), an instrument especially designed for this study. This measure was administered before any interaction took place, to insure that the responses reflected the notions about the content and purpose of evaluation held by the teachers before they had been exposed to the feedback intervention.

Three separate types of agendas had been arranged, one for each of the three treatment groups. The meetings were basically the same except that each site was informed as to which treatment group their classes had been assigned and given a brief description of the total evaluation package for their own as well as the other treatment groups. After the program for the year was presented, the Data Aide was introduced and her function explained. The meetings were then thrown open to discussion and questions.

The following is a summary of the types of questions and comments at these orientation meetings:

1. Requests for Specific Information:

    a. Who will be doing the testing and observation and what are the qualifications?

    b. Will teachers in the research study be given help or materials for the on-going program?

    c. What are some of the findings from research studies carried out in the Head Start Centers during the past two years? For instance, would the study on the use of dialect vs. standard English lead to down-grading the local speech?

   d. Would the UCLA representative explain the program at a parent meeting?

2. General Information:

   a. Will performance on the tests affect the future of the Head Start program?

   b. Will the results have any effect on the teachers' ratings or employment?

   c. If there is resistance to home interviews, because of previous unpleasant privacy-infringing experiences with welfare workers, how will this be handled?

3. Criticism of Evaluation Instruments:

   a. Validity of Binet and other tests which have a middle-class bias;

   b. Ambiguity of verbal stimuli in the Gumpgookies test and the difficulty and irrelevance of its concepts;

   c. Administration of tests in English to Spanish-speaking children or to those children whose first language is not English but who speak just well enough to take the test;

   d. Aversion to being videotaped and general objection to participation in the program;

   e. Objections to filling out teacher instruments, particularly the TATE;

   f. Failure to get any meaningful information about children from the previous year's evaluations.

At the end of the meetings the teachers were asked to fill out the Teacher Expectations for Achievement of Children in Head Start (TEACH).

III. Feedback Meetings

Feedback Session #1:

Test results for each sample child were given in general descriptive terms, with emphasis on those areas where teachers could, if they chose, utilize the information immediately in their work with individual children. Summaries of test results were used to explain the purpose of the various tests. Note that no scores were given, only a statement as to how adequately the child was functioning in each area. For example, the report might say: "(child's name) showed a very irregular pattern of performance, sometimes failing an easy question and doing very well on a harder one in the same category. He expresses himself well, but often his statements are not about the subject under discussion. This type of behavior may be due to lack of concentration. He needs help in developing his attention span and visual motor skills."

In most cases the teachers agreed that the test results accurately described the child's behavior as they saw it, and they seemed pleased to have confirmation from an objective source. However, in those instances where the data contradicted their own impressions, two types of responses were likely: the first was an acceptance of the test findings and an attempt to develop remedial activities for the child; the second was a rejection of the low scores as not truly representative of the child's ability. In the latter case, either the tests used or the child's lack of rapport with the strange examiner in the unfamiliar testing situation, or both, were considered culpable.

Disagreement between test results and the teacher's opinions seemed to occur primarily in the area of verbal expression. Either the teacher found that a child who expressed himself freely in the classroom was shy with the tester, or a child who was very quiet in the group talked a great deal in the one-to-one situation with the tester. In some cases, teachers asked for information about certain techniques they had heard about and were thinking of using to overcome the child's difficulty. In keeping with the limitations imposed on the feedback intervention, no prescriptions were supplied, although Socratic questioning encouraged free discussion. At the sites where the Child Development Supervisor attended the feedback meetings, she often picked up ideas from these discussions and made suggestions as to ways the teachers could handle the specific problem.

At this first feedback meeting, the teachers were given copies of a paper entitled, "The Use of Games with Preschool Children to Increase Perceptual, Conceptual, and Verbal Skills" prepared for Head Start in Summer 1966 by June Mayne and Ruth Pearce. It was hoped that teachers might be able to adapt some of the ideas presented in this article since the cognitive skills are the ones in which Head Start children are usually most deficient, but there was no discussion of the material itself or how it might be used.

In preparation for the next feedback session, where the videotapes were to be viewed, a brief explanation (see page 56) of the observation procedures used in the national evaluation, and sample copies of the classroom observation instrument, the OSCI, together with the scoring sheets and coding system, were distributed.

Feedback Session #2:

At this meeting the first series of videotapes was shown. Two classes met at a time so that teachers could see their own class as well as one other, for comparison purposes. Before the tape was started, the teachers were asked what they hoped to see and in what way the information might be helpful. Typically, the teachers wanted to see a) the interactions among their children which they might have missed because of their involvement elsewhere and b) how effective they were as teachers. The teachers had been informed that they would be asked for comments and feedback as to whether the tapes were a fair representation, whether they were able to pick up situations or behavior they might otherwise have missed, and whether the tapes could be of any use to them as teachers.

The initial reaction from the teachers was characterized by a great deal of self-consciousness about their personal appearance. They admitted that they had not been as relaxed as usual while the filming was going on. One teacher said that it had not been a typical day for her in that it had been rainy and the children had been confined indoors all morning and were unusually restless. The Child Development Supervisor present at the meeting suggested that the children be broken down into smaller groups on rainy days. Another teacher who had also claimed that the videotaping had not been done on a typical day, after viewing the tape was surprised to find that it was much more representative than she had expected. This same teacher used the tape as a means of evaluating the work of the other adults in her classroom. She felt that she and her assistant functioned well but voiced concern because the tape showed that the Neighborhood Youth Corps worker was not clear as to her role, and was not functioning as well as she might if she were given more specific direction.

One teacher felt that the tape showed how vaired were her activities in a single morning. She commented that the teacher and assistant teacher must constantly assert themselves as focal people in the class due to the many adults who pass in and out of the room during the day, such as personnel from other classes, the nutritionist, nurse, etc.

One teacher's first reaction was that everything was beautiful. After a good deal of discussion she said, "Everything was good, except for me." When asked what she meant, she went on to say she didn't look relaxed on the tape. The Feedback Analyst commented that it is difficult to relax while being filmed, but the teacher said she had not known that the camera was focused on her at the time. She remained quite concerned about the way she looked.

Several teachers were concerned about seeing so few instances of teachers working with individual children. Discussion confirmed the need for such personal interactions, but it was generally agreed that the quality of the contact, even those of short duration, was far more important than the frequency with which they occurred.

Two teachers were very pleased with the impression they conveyed. One said, "I never knew we were so great." These two teachers used most of the ensuing discussion to talk about individual children seen on the tape. Many teachers requested that the tapes be presented at a parent meeting.

When asked what kinds of events they would like to see on future videotapes, all teachers expressed a desire to view tapes of other classes and teachers as a way to learn different techniques and teaching styles. They wanted to see episodes involving more constructive involvement of assistant teachers, creative use of free play periods, doll corner play, and teacher-structured cognitive learning situations. They also wanted to see how "good" teachers handled "acting out" and "quiet" children.

Feedback Session #2:

The Feedback team planned to bring the Parent Interview form to a feedback meeting for discussion before using it in the field. This was not possible because the protocols did not arrive from Washington until long after they were due, and had to be taken out immediately to collect data within the pretest period. Thus the third feedback meeting came after most of the Parent Interviews had been completed. It was therefore decided to present the form along with some of the types of information gathered, especially that which related to the children in the class. Discussions centered around the attitudes of parents toward the Head Start program, how far they expected their children to go in school and how they would make out in life generally, how they prepared their children for the first day in school, how they used punishment or rewards, what they knew about community resources and services, and their opinions about life in general.

The Parent Interview revealed that most parents have a pretty good idea of what Head Start hopes to accomplish and a positive attitude toward the program. Many parents wanted their children to go to college and pursue professional careers, but their hopes of ever achieving these aspirations were quite limited; except for one parent, the highest level of education actually expected was high school graduation. When asked what they told children on their first day at school, most mothers indicated that their primary concern was for the child to conform to the teacher's expectations. Typical responses were: "Do what the teacher says," "Don't fight," "Sit in your seat," etc. While it is quite possible that the parents were giving the responses they thought the interviewer would find most acceptable, the emphasis on social conformity seems to reflect realistically the parental attitudes of this group.

With respect to child rearing practices, most parents revealed lack of knowledge of age appropriate behavior, with little relationship between the child's misdeeds and the parentally-imposed discipline. Spanking was still the most common type of punishment. Although many parents knew about community medical services, few of them were aware of where to go for legal help, education or vocational advice. On the questions reflecting feelings of powerlessness and frustration, most parents expressed a high degree of alienation. However, the teachers felt this might have reflected a fatigue effect because these items were at the end of the extremely long interview.

The teachers were very interested in the responses to the questions on child rearing practices and felt that there were many openings for parent-teacher discussions. It was also felt that the transition into elementary school needed more exploration in parent education sessions.

Feedback Session #4:

The purpose of the fourth feedback meeting, held during the month of February, was to allow teachers to view the second series of videotapes, but because the video camera had been broken by one of the Head Start teachers knocking over the tripod, the film showing at all but two sites had to be postponed. The content of this meeting was therefore devoted to a description and discussion of the Picture Playboard Sociometric Test which had been administered in the mid-period months of December and January. The Data Aides demonstrated how the instrument was administered and then the teachers were given a record of each child's choices. The Feedback Analyst invited the following types of comments: (1) comparison of the child's social level when he entered the program to his present level of interaction; (2) judgments as to whether the child's choice of playmates on the sociometric instrument was a true reflection of observed interrelationships in the classroom; (3) speculation about the meaning of choices which seemed to be contrary to the teachers' own observations; and (4) ideas as to how the sociometric material could be utilized by the teachers in their work with specific children.

There was considerable variation in the amount of comparative material that teachers were willing or able to provide for each child under discussion. Comments ranged from, "He cried a lot when he first started, but doesn't any more" to "When she first came in she was dominated by her cousin and did not play with the other children. Now she is playing with other children and can stand up to her cousin."

When the actual sociometric data was presented, there was a wide range of responses. In one class, the teachers found the reports paralleled their own perceptions of the children and made only limited remarks such as, "He's not very social" or "He gets along well with everyone." However, most of the teachers were surprised at some of the children's choices. In speculating about this incongruence, at least two teachers from different sites suggested that these unexpected choices were actually a reflection of the desire to play with a certain child rather than an actual friendship. They noted that many of these projective choices were by shy, withdrawn children for aggressive, out-going children. In the cases where teachers did not spontaneously offer projection as an explanation, the Feedback Analyst offered this as a suggestion and the teachers, in most cases, agreed that this was an interesting possibility and that this information could be useful in bringing certain children together. In the ensuing discussion, one teacher noted that the same mechanism might be operative when children who were unable to form desired associations resorted to hitting and other forms of hostility. A teacher said she would use the results as a guide in suggesting to certain children who they might like to be near or play with. In one class, a discussion came up about the most chosen children and their emergence as possible leaders. One teacher intimated that in her class the leaders were usually the disruptive children.

In trying to understand why certain children were not chosen, it was pointed out that one was a child whose mother is almost always present in the classroom as a volunteer; this child is actually more animated when the mother is absent. Another unchosen child was said to come from a family where there was a great deal of physical violence. He talks in school about monsters and likes to play "baby" which the other children permit. Another child does not talk, but likes to look at books. It was suggested that the teacher might use books with this child in a one-to-one situation, to establish communication and rapport. Although the sociometric instrument is in a very experimental stage, and its reliability and validity are still very much in question, it was found to be an effective springboard for eliciting discussion about children's peer relations. In general, the teachers felt that the feedback from this instrument would help them become more observant and aware of children's interaction patterns.

The teachers who viewed the videotapes after the discussion of the sociometric measure were so engrossed in observing whom the children selected as their playmates that they forgot to be conscious of their own appearance. At one site the Child Development Supervisor questioned the value of an art activity which involved tracing forms. This elicited a discussion among the teachers in which they defended the use of structure in fostering creativity. When the teachers were asked to indicate whether or not they thought the video tapes were useful, they responded affirmatively; they felt that the tapes enabled them to see activities and interactions of which they might not normally be aware. One teacher, seeing a child sitting and stringing beads alone for a long time, regretted that she had not been aware of this so that she could have steered the child to another activity. In another case, a teacher noted a child whom she had thought to be hyperactive, with a very short attention span, sitting and quietly painting for an extended period of time.

Feedback Session #5:

A separate agenda was planned for the two sites which had viewed the videotapes at the fourth feedback session. At one site the teachers chose to discuss the Post Parent Interview and the Post Program Interview. The reaction to these instruments was generally favorable. The only negative criticism was addressed to the item on the Parent Interview which asked for the source of the family income. Most teachers felt this was an unnecessary invasion of privacy.

At the second site, the teachers asked about the tasks related to the TEACH instrument. Because of this interest, they were invited to participate in the pilot work, suggesting items as well as appropriate objectives. For the other two sites the fifth session was devoted to jointly viewing the videotapes of their classes. Because there had been some criticism of the quality of the sound on the videotapes, at Site #3, where roughly comparable art activities had been recorded, one tape was played with sound and one without. Both Head Teachers preferred the tape with sound while the assistant preferred no sound. It was interesting to see that the art activity showed children coloring in pre-cut, pre-pasted figures and neither the teachers nor the Child Development Supervisor commented on the uncreative nature of the task. At the fourth site, after viewing the first tape the Head Teacher of the class involved expressed some concern about the children's noisy and unruly behavior. In the discussion, it was suggested that some of this disorder might be a reaction to the more relaxed atmosphere being fostered by a new Head Teacher who had replaced a very rigid disciplinarian. The new teacher said that she was making an attempt to introduce a more appropriate learning environment by adjusting the physical arrangements in the classroom. One of the techniques she was using to involve children was to invite participation in her work on the flannel board.

During the showing of the videotape of the second class at this site, the Head Teacher recalled a conversation about a particular child which grew out of the discussion of the Playboard Sociometric at Feedback Meetin #4. The teacher called attention to a child previously docile and subservient, who was now holding her own on the videotaped segment. The responses of other children to an emotionally upset child ("He cries, we don't cry. He's not ready for the big school") were discussed, with the suggestion that these comments revealed their own anxieties about "big school."

In all of the sites, at the end of Feedback Session #5 the teachers were asked to suggest areas for future investigation by the Head Start Evaluation and Research Center. Several teachers indicated interest in a follow-up study of the children currently in their classes. They felt this should include measures of both academic and social progress in the school as well as at home and in the community. Another suggestion was the setting up of parent meetings where child-rearing techniques could be taught. Further proposals along similar action-oriented lines indicated either a misunderstanding of the nature of basic research or disinterest in this type of endeavor compared to goal-directed action. Probably a little of both is reflected in the impatience of the ghetto communities at the lack of perceived relationship between studies carried out by researchers and establishment of programs to alleviate their problems.

Feedback Session #6:

Because of the interest expressed in the organization and activities of the UCLA Evaluation and Research Center by a number of teachers, an extra feedback session was planned for the 12 experimental classes. While a large group turned out for the meeting, the attendance was far from complete. One site with five classes, which had been in favor of the meeting, was represented by only one Head Teacher and no assistants. To the surprise and consternation of the Feedback staff, the Head Teacher of one of the Control classes appeared as the substitute Child Development Supervisor. It may very well be that this type of interaction between experimental and control groups may have confounded the treatment effects.

The program for the afternoon meeting included a period for socializing to allow the teachers and the Center staff to get to know one another, a tour of the Center facilities, with special emphasis on the data processing and production departments, and a film. The film was one produced by Dr. Marianne Marschak and was concerned with mother-child interactions in a task situation. The film showed four Head Start mother-son pairs involved in several pre-set tasks. The episodes revealed that there were four basic patterns or ways in which each mother worked with her child: (1) the mother who intervened in the task only when her child obviously needed help; (2) the mother who constantly directed and controlled the child's activities; (3) the mother whose own need for play was so strong that she became completely engrossed in manipulating the materials to the exclusion of her own child; and (4) the mother who was uninvolved and disinterested in the tasks to the point of ignoring the child's cues for help.

Following each diadic presentation there was a short sequence in which the child was filmed while the mother was out of the room. Some children were shown to be restless and unhappy while others worked on their tasks contentedly by themselves. However, no overt inferences were drawn relating this behavior to the type of mother-child interaction which had been observed. It was felt that showing this film would not only make teachers more aware of different styles of mother-child interactions, but also make them aware of their own ways of behaving with their children.

Dr. Marschak, the maker of the film, was present during the showing of the film and was available to answer questions. However, only three or four teachers took an active part in the discussion period. The others seemed either tired or disinterested and a number of them were seen to doze off in the darkened room. Those who were active in the discussion seemed to agree that it might be an excellent procedure to observe mother-child interactions at the outset of a Head Start program, to provide teachers with insight into individual parent and child behavior patterns. This technique was also seen as a way of introducing mothers to Head Start materials. None of the teachers reported any emphatic identification with the mother-role in the film.

Final Meeting

Late in May, 1969, the teachers of the experimental and control groups met at their individual sites with representatives of the UCLA Evaluation staff. The experimental groups met with the Feedback Analyst, Assistant Feedback Analyst, and the Data Aide assigned to the site. The Evaluation Coordinator, a Classroom Observer, and the Data Aide attended the meetings at the Control sites.

The three major components of the final session included: 1) presentation of gains made by individual children on the Binet, the Caldwell and the Playboard Sociometric; 2) posttest administration of TATE and TEACH; and 3) discussion of teachers' reactions to the feedback and evaluation program. In the case of the feedback classes, teachers were asked whether the feedback program had been instrumental in improving their attitudes toward evaluation, whether the feedback from test performance had helped them understand the needs of individual children, and whether this information had made them better teachers. The teachers in the Control groups were asked how they now felt about evaluation and whether they resented the fact that they had not been given the type of feedback the experimental groups had received. Discussion also centered about the expressed attitudes of teachers toward the national evaluation program, its effect upon them and their children, and what they felt they had gotten out of it.

Feedback at Experimental Sites

The results of the posttests for each child were presented in a summary form and the teachers then asked how the data matched their own evaluations. One teacher confessed that she had had very low expectations of improvement for her class. As a whole they were the most difficult group she had ever had, were easily distracted, and had very poor attention span. She felt that the site telephone, which rang a great deal, created an important source of interference since it was located in her area.

Another teacher commented that the children who showed the least amount of improvement were those with very serious problems or those whose attendance had been particularly poor. In one case where no real improvement was noted, the teacher felt this was directly due to an over protective mother. This comment led another teacher to report that one of her mothers, who had resisted sending her child to school, became involved in the Head Start class activities and was now volunteering on a regular basis. This teacher also mentioned that many parents who had become involved in Head Start were stimulated to return to school to further their own education. A teacher who had kept rather detailed anecdotal records made the interesting observation that the most dramatic changes took place about the third week of the program. However, she was referring to such affective areas as relaxation of tension and increased social interaction.

The reactions to the posttest data from the Playboard Sociometric were very perceptive in pointing up several serious weaknesses in this instrument. For instance, it was noted that a child's choice of a playmate may be inconsistent from day to day, and reflect a transitory mood. Also, young children are very fluid in their choice of a playmate and their choices reveal the preference for an activity rather than the particular child selected. It was discovered quite by accident that a child who was chosen by several other children had just had a birthday party at the site the day the sociometric test was administered. These types of events often went undetected, yet they certainly had a confounding effect. While the test was thus an inadequate measure of change, it did reveal some interesting information. For instance, an active, outgoing child consistently selected a shy, quiet child who cried a great deal. Several interpretations of this behavior were advanced and the one which seemed to satisfy the teachers the best was the comment that the child was empathizing with her unhappy peer.

In general the teachers tended to restrict their comments to very specific problems and behaviors, such as concern about two children who were suffering from serious emotional and learning problems. It seemed quite clear that these children would not be able to handle the regular school situation. The Child Development Supervisor solicited the assistance of the evaluation staff in obtaining special help. A similar instance of attention to a specific situation was brought up by a teacher whose student scored low in a test of visual discrimination. The unexpected incongruity between the child's test performance and her own knowledge of the child led the teacher to request an eye examination, with the subsequent diagnosis that the child needed glasses. After corrective lenses had been provided, the child's visual skills were no longer out of line with her other performance levels.

The second component of the final session was the filling out of the TATE and TEACH scales by all teachers. While the post form of the TATE was the same as the pretest, on the TEACH, the teachers were first asked to rate the behavioral objectives according to degree of importance on a ten point scale, and then to indicate to what extent they felt they had succeeded in achieving these goals. Because the presentation of the results had taken so long, the teachers were not at all happy about spending additional time completing these forms. As a compromise, they were asked to complete only the TATE, the shorter instrument, at the site and were permitted to complete the TEACH at home but cautioned to be sure to return it the following day with the Data Aide. Even so, there was considerable objection to the TATE, with several teachers refusing to answer the items seriously.

Reaction to the Feedback Program (Experimental Sites):

The reaction of the teachers in the experimental group to the feedback program was, with few exceptions, favorable. Even at the site with the most negative teachers, there was progress from outright hostility at the beginning of the program to tolerant indifference at the last meeting. In general, the best teachers were most receptive to and enthusiastic about the program. A dramatic positive change was seen in a teacher who went from extreme suspicion and hostility, which she displayed disruptively at the Orientation Meeting, to a strong positive feeling about the evaluation staff and the value of the feedback material. Another teacher expressed her appreciation for the data which had been provided to her during the intervention program. She said that the feedback had helped her in preparing materials appropriate to the needs of individual children. While the teachers could see the value of the tests with the children, they repeatedly questioned the value of the various teacher measures such as the TATE and the TEACH, which took up teacher time.

The tenuous nature of the rapport with the Head Start staff was pointed up by an unfortunate event in the week immediately preceding the final meeting. The Feedback Team arrived at a site to find all the teachers terrible upset by the visit of a researcher who had asked them to rate the children in their class from most likely to least likely to succeed. The young man had asked the teachers to perform this arduous task while the class was in session, contrary to the basic policy of the E & R Center not to make any demands on the teachers during class time. It was only with the greatest difficulty that the teachers could be convinced that this research was neither part of nor authorized by the Center, but was actually a study a graduate student was doing for a doctoral dissertation.

Final Meeting: (Control Group #1)

The reception given to the evaluation staff at the control sites varied from positive warmth to open hostility. The latter attitude characterized one teacher who felt that she had been forced by her agency to participate in the evaluation program against her wishes. She had been cool and passively uncooperative with the evaluation staff during the program; however, at the final meeting she confessed contritely to the Aide assigned to her site that she had set out to give her a bad time and had used her as the scapegoat for her hostility toward the "establishment." It is quite unlikely that the antagonism had been engendered by a fear that evaluation would show her to be a poor teacher, since all the observers reported her to be extremely competent. Most of the Control group teachers did resent being excluded from the feedback program; they felt that the information about the children which they were getting at this final meeting could have been very helpful during the year but was completely meaningless now that the children were about to leave the program. Some teachers were so concerned with specific items on the tests that it seemed as if they regretted not having had the protocols as a guideline or syllabus on which to base their instructional program. Most teachers wanted to know how their classes compared with other Head Start classes. The evaluation staff attempted to avoid comparisons among teachers and emphasized the need to focus on changes in children.

While most teachers accepted as valid data which demonstrated growth in either cognitive or social-emotional areas, they challenged the test scores when no measurable changes were obtained. The most common rationalization was the inappropriateness of the tests for this population. A second criticism was that the testing situation was unfair and that the children actually knew more than they had demonstrated with an unfamiliar examiner. Many teachers felt that if they had tested the children the scores would have been much higher. A third type of explanation for children's failure to improve related to serious problems in the home and family situation. Very few teachers were willing to admit that poor learning might be related to poor teaching.

There seemed to be a discrepancy between the treatment of the Data Aide and that of either the Evaluation Coordinator or the Feedback Analyst. For example, when the Data Aide attempted to make after school appointments to complete the required national evaluation instruments, she was repeatedly told to come back at another time. At one site this procrastination and passive resistance was so frustrating to the Data Aide that she asked the Coordinator to come out to the site and obtain the data for the Post Program Interview with the Teacher.

Again, much resistance to filling out the TEACH and the TATE was evident. While a good part of this was due to the heavy burden of paperwork which was their regular responsibility, there is no doubt but that their inability to see the relevance or importance of these instruments constituted a major basis for their protest. As with the experimental groups, the teachers were allowed to complete the forms at home. When asked for suggestions as to future research or evaluation activities, many of the teachers wanted to be included in the feedback group.

Final Meeting (Control Group #2):

A supplementary hypothesis of the intervention was that the teachers of classes receiving the total evaluation package, with its heavy load of individual testing, classroom observation, and teacher questionnaires, would even without the feedback feature of the experimental treatment, demonstrate measurable changes in their attitudes toward evaluation. The constant interference which is an unavoidable concomitant of evaluation did produce some degree of strain, but for the most part the relationships between Head Start and evaluation personnel remained cordial. In spite of the requirement in controlled research studies to maintain an attitude of impartiality and objectivity, a warm rapport was established and in many cases the evaluators were made welcome even when the task they were there to perform was onerous.

To counterbalance in some part the effect of these recurrent contacts with evaluation personnel, a second control group, which received only a minimal test program, had been included in the design. However, even this minimal package required the presence of testers in the classroom for several weeks at the beginning and end of the program. In addition, the Data Aide who was assigned the task of keeping attendance records and filling out the various classroom inventories, found it necessary to visit these sites far more than had been anticipated. At one site the teachers were consistently shorthanded and the Data Aide was called upon to help out every time she appeared. She was in a difficult position: if she refused to help she would create hostility and resentment, and if she helped she might introduce a confounding element into the design. In taking the easier alternative of assisting the teacher, she succeeded in establishing a strongly positive relationship which undoubtedly influenced the responses on the posttest teacher attitude measures.

The final meeting for Control #2 began with a review of the evaluation intervention design and an explanation of the role played by this group, which had received only a minimal test battery and no classroom observation. However, the teachers were not willing to accept the reasons given for their being excluded from the feedback program.

When the scores on the tests given the individual children were discussed, low gains were again attributed to home problems, shyness, lack of examiner rapport, and the inappropriateness of the test instruments themselves. Reaction to the evaluation was generally either an outspoken criticism or a passive acceptance of the inevitable. When asked for constructive alternatives, given the legal requirements for evaluation, a number of concrete suggestions were offered. One approach which had merit although probably presenting many logistic problems, was to film children at various points during the Head Start year. Such a film would be able to document dramatic changes, for instance, a shy, withdrawn boy emerges as a happy, friendly child. This type of record might be far more effective than the most sophisticated statistical report.

A complete lack of understanding of researchers and test construction was evident in the suggestion that experienced classroom teachers should design the tests because "researchers were never around children." There was general concern among teachers over lack of feedback, tests not giving a true picture of a child's capabilities, no way of measuring social gains and the impact of Head Start on the family and the community. Although the Playboard Sociometric was not used with these groups, it was described and many teachers who were critical of the cognitive tests would have liked to have this instrument administered in their classes.

Supplementary Contacts with Agencies

Aside from the regularly scheduled meetings, agencies and/or teachers utilized the services of the Evaluation and Research Center in various ways. While in some cases requests from agency personnel were perceived as an attempt to allay their own suspicions and clarify for themselves just what the evaluation was about, in many instances the Center was used as a resource through which they could supplement their own programs. A brief summary of contacts with the four Delegate Agencies follows:

Agency I (2 classes). The Training Supervisor requested a preview of the first videotapes, before meeting with the teachers. The ostensible reason given was to see if the tapes contained any material which might prove threatening to the teachers. A meeting was held at the agency headquarters with the Training Supervisor, the Project Director and two Child Development Supervisors. After the viewing, they asked if they could have copies of the tape for teacher training purposes. The Child Development Supervisor for this particular site attended two feedback meetings. Teachers at this site asked if the video tapes could be used at a parents' meeting. The Training Supervisor for this agency also called the Evaluation and Research Center office and requested two copies of the report on the use of games with preschool children, which had been given out at the Feedback meeting.

Agency II (4 classes). The Training Supervisor was present at Feedback Meeting #1, at which time test findings were discussed. She had many questions about the testing materials and a special meeting was arranged with the Feedback Analyst and the Evaluation Coordinator. At first the Training Supervisor challenged the instruments on the basis of their middle class bias, but as the meeting progressed and the manner in which the measures were used was explained, it became apparent that the supervisor was really upset at the requirement from the local office that the teachers administer the Preschool Inventory. During this meeting it was learned that this agency had adopted the procedure of asking teachers for observational and background material for all their children and were devising a system of getting this material from all teachers in their agency. The Project Director of this agency attended the third feedback meeting.

Agency III (2 classes). The Child Development Supervisor attended Feedback Meetings #1 and #2 at this site. She entered into discussion and used the material presented at these meetings for teacher training purposes. The teachers at this site scheduled a parents' meeting just preceding Feedback #1 so that the Parent Interview could be clarified. The teachers requested the attendance of the Feedback Analyst at an evening parents' meeting in order to show the video tapes. At Christmas time the Head Teacher reported that they did not have individual gifts for their children. The Evaluation and Research Center staff personally contributed enough so that each child could get a present.

Agency IV (2 classes). The maximum participation was experienced with this agency. The teachers were extremely cooperative and very candid in feedback discussions. The Child Development Supervisor attended all feedback meetings. In addition, the Training Supervisor, the second Child Development Supervisor and the social worker attended the first feedback meeting.

The personnel at this site requested that the video tapes be shown at a special parents' meeting. In addition to the Child Development Supervisor, the Training Supervisor, social worker, and Project Director attended. At this meeting the Feedback Analyst talked about the general design of the feedback study and suggested that they ask questions after viewing the tape. The videotape covered a range of activities, including a singing game, snack, cleanup, and two table activities, and showed both Head Teacher and Assistant equally involved with individual children. Most of the teacher-child interactions were concerned with either cognitive learning or the exploration of experiences and feelings. After the tape was over the teacher appeared relaxed and pleased but ended the meeting without a question period, saying, "I don't have any questions and I don't think the mothers do after that lovely presentation."

The teachers asked if they could see video tapes of other experimental classes, and if they could arrange to visit other nursery schools. A meeting was set up at the Evaluation and Research Center office for the purpose of showing videotapes of other experimental classes for these teachers. The Child Development Supervisor invited all the other teachers under her jurisdiction to this meeting and conducted a training session after the videotape had been shown.

There was no question but that, subjectively at least, the Feedback intervention was successful in reducing antagonism to evaluation and increasing rapport with teachers and agency personnel. Such results, even if not measurable in as quantifiable context, are certainly worthwhile from a human point of view. It is recommended that in future research and evaluation efforts, teachers be informed of the objectives and outcomes of the work carried out in their classrooms.

DESCRIPTION OF NATIONAL HEAD START COMMON CORE MEASURES

I. Individual Child Measures (Pre- Post; October and May)

1. Stanford-Binet Intelligence Scale.
2. Birch Response Style. This is an interesting new way of recording the manner in which a child responds to the items on the Stanford-Binet Scale. Instead of simply scoring items pass or fail, the examiner records verbal responses, verbal elaborations, non-verbal responses, refusals, etc.
3. Revised Preschool Inventory. A new standardized version of an instrument which measures achievement.
4. Gumpgookies Test of Achievement Motivation. A two-choice picture test developed at the University of Hawaii.
5. Animal House of Wechsler Preschool Scale of Intelligence. A short measure of cognitive ability.
6. Sociometric Picture Board. A measure of child's social adjustment. Children select, from poloroid pictures of their classmates, the child they wish to play with in several structured situations.
7. Medical Data. Obtained in collaboration with agencies contracting to do physical examinations with each Delegate Agency. Hemocritic blood tests, height, and weight data.

II. United States Bureau of Census Descriptive Data of Head Start Centers.

(Evaluation Center will gather this data on sample classes only, and return it to the Census Bureau.)

III. Parent Interview. Conducted with parents of sample children in October or November and again in April or May.

IV. Descriptive Data on Sample Classes. These include attendance records, classroom facilities inventory, characteristics of teaching staff, class record forms, etc. This information will be gathered by Data Aides.

V. Classroom Observations. Observers visit each class about once a month to record ongoing curricular activities of individuals and groups within the class. Emphasis is on the activities of the children rather than the behavior of teachers.

INFORMATION ABOUT CLASSROOM OBSERVATIONS

PRESENTED TO TEACHERS IN ORIENTATION MEETING #2

There will be four Classroom Observations, one each during the months of November, January, March,and May. These observations record the different kinds of activities which occur over a two hour period. The observations of the class includes the type of activity, the materials in use, the number of children involved in the activity, and the person in control (child or teacher) at a given moment. There is also a record of input from the teacher and socializing among children or between children and teacher.

The description of the activity relates to such questions as whether the child is refining his listening skills, is involved in using large muscles, or learning a new skill. Is he engaged in an activity which involves visual discrimination, verbal communication, etc.; or is he repeating and mastering a familiar activity?

Children and teachers are observed in alternate segments. In the teacher observations the number of children with whom the teacher is involved, their activity, and her input to them, is noted. The observations are in three minute segments, broken into 20 second records. The assistant teacher is observed in the same manner.

It is hoped that this observation instrument or procedure will reveal the wide variety of Head Start classrooms across the country, or within regions; the effect of different types of classroom environment on the children's progress will then be determined.

CATEGORIES OF SKILLS ON STANFORD-BINET AND ITEMS RELATED TO THESE SKILLS AT VARIOUS AGE LEVELS

I. Visual-Motor

Tasks involving eye-hand coordination, e.g. drawing, folding papers, stringing beads (II 1, 2, 4; III 1, 3, 5, 6; III 1/2 2, 5; V 1, 2, 4, 6; VI 6; VII 3).

II. Verbal

Tasks such as labeling and word recognition. A more advanced level in this category requires definitions of simple words, verbal explanations of differences between objects or animals, and comparison of sentences containing similar elements. E.g. Food is to eat, water is to _____ (drink). (The Binet has been criticized as being too heavily weighted in this area.) (II 3, 5, 6; II 1/2, 2, 3, 4; III 2; III 1/2 4; IV 1/2 2; V 3; VI 1, 2, 5; VII 2, 5).

III. Memory

Tasks that require short term memory of numbers, pictures, and objects. The evidence can be recognition, reproduction, etc. (II 1/2 5; III 4; IV 2; VII 6).

IV. Following Directions

Tasks that require the execution of simple commands, such as, "open the door", "put the button in the box", etc. They require comprehension of spoken English and familiarity with the objects involved, as well as some memory. (II 1/2 6; IV 1/2 5).

V. Visual Discrimination

Tasks that require recognition of similarities and differences between pictures. At a more advanced level, identification of missing parts in familiar objects, or of parts that do not go together. E. g. Shown a picture of four cups and a saucer, the child is asked to point to the one that is different. (III 1/2 1, 3; IV 4, 5; IV 1/2 1, 3; V 5; VI 3; VII 1).

VI. Comprehension

Questions that require familiarity with common objects and adequate verbal mastery of the language to express this familiarity. E.g., "Why do we have chairs?" (III 1/2 6; IV 6; IV 1/2 6; VII 4).

VII. Numbers

Only one task, at level VI, to determine ability to count objects up to ten. (VI 4).

VIII. Information

Questions that require a knowledge of the use and composition of simple objects. E.g., "What is a table made of?" (II 1/2 1; IV 1/2 4).

EXAMPLES OF ITEMS ON SUBTESTS OF PRESCHOOL INVENTORY USED IN FEEDBACK REPORTS

A. Self Identity
   What is your first name?
   What do you call this (point to elbow)?

B. Language
   Say, "Hello", very loudly.

C. Color and Position
   Put the yellow car on the little box.

D. Understanding and Following Directions
   Give everything to me.

E. Knowledge of Community
   Who do you go to when sick?
   Where do you buy gas?

F. Numerical Concepts
   How many eyes?
   Count to 5.

G. Quantitative Comparisons
   Which is bigger - a tree or a flower?

H. Position Concepts
   Point to the middle checker.

I. Abstractions
   Which is most like a wheel?

J. Visual-Motor
   Copy this line.

K. Color
   What color is this (black crayon)?

L. Shapes
   Color: Circle, square, triangle

THE DATA AIDE PROGRAM

A major problem in carrying out evaluation with minority groups has been that of finding qualified personnel of the same ethnic population to administer tests and conduct home visits and interviews. If, as is usually the case, the basic criterion for qualification is extensive professional training at the college level, the pool of potential evaluators of minority background is indeed very limited. Another problem associated with evaluation is the tremendous load of paperwork which falls to the lot of the teacher caught up in a sample evaluation class. In addition to her regular records, she must fill in various types of inventories and protocols to describe such details as teaching staff, classroom facilities, program activities, parent participation, etc. It is certainly no wonder that teachers balk at being involved in evaluation.

In an attempt to speak to both of these problems, the UCLA Head Start Center decided to employ paraprofessional Data Aides recruited from the local communities, to be trained to keep records, fill in forms, and administer the less demanding individual child tests. Community Action Programs involved in skill-training of indigenous personnel were invited to make referrals for the Data Aide position. The flyer used in recruiting is shown on page 62.

All four data aides recruited were residents of the areas served by Head Start. One had worked as a Nutrition Aide in a Title I Preschool Program; one had worked as a bookkeeper and was attending evening classes at a City College to fulfill the course requirements for becoming a preschool teacher; one had worked as a volunteer in a Head Start class which her own children attended; and the fourth recruit had been a day worker most of her life and had just completed a community-sponsored skills training program in the operation of duplicating machines. All four Data Aides had expressed a strong desire to work with children and to improve their job skills in order to become upwardly mobile.

The Data Aides were trained right along with the more highly qualified staff but were given extra help and assistance in mastering the skills of testing and data collection. The training for administering the individual child measures consisted of a period in which to become familiar with the manuals and test materials; a period of administering the tests to one another under supervision; and finally, practice administration on a few children. Teaching aids such as tape recordings and video tapes were used to help improve voice quality and to create an awareness of the importance of facial expression in establishing rapport with children. The initial reaction of the Aides to the tapes and films was one of shock and self-consciousness, but these media served as excellent tools both in improving the aides testing skills and in developing their self confidence. Data Aides were given special instructions in the gathering of descriptive data such as daily attendance records and class rosters. They were also asked to keep a verbatin log of the happenings at their sites during the times they were in attendance, recording such features as drastic changes of schedule, visitors to the site, trips, absences of teachers, etc. The primary purpose for keeping these logs was to be able to identify any unusual occurrences which might bias the results of the evaluation study.

The Parent Interview training was conducted by a special training team from Washington. The Data Aides participated in this two-day training along with the other interviewers, almost all of whom were students at UCLA with at least two years of college. The Evaluation Coordinator attended these training sessions and concurred with the Washington trainers that the Data Aides needed very little extra training and completed the sessions at least as well-qualified as the more educationally-sophisticated trainees.

On the whole it was felt that the Data Aide program was very successful. In almost all cases, the Aides were cordially welcomed by Head Start staffs. They had excellent rapport with both staff and children, and their presence in the classroom was generally appreciated, since they were available to help the teachers when they were not busy with evaluation tasks. At the final meetings, most teachers mentioned the Data Aide as one of the best features of the evaluation program. One teacher felt that it had been unfair to these paraprofessionals to ask them to serve as the liaison between the home and the school "establishment". The same teacher now expressed remorse over the hard time she had given the Aide assigned to her site. There was only one site which did not cooperate fully with the Data Aide in providing access to the required information. The teachers at this site always complained that they were busy and asked the Aide to return at another time. The Aide finally had to call upon the Coordinator for help, and some pressure had to be brought to bear to obtain the necessary data. These were the only negative incidences recorded, compared to many positive ones.

The program proved that paraprofessionals certainly can do an adequate job of administering tests to Head Start children. If there were any one particular problem, it was the tendency to adhere only very loosely to the standardized procedures which must be followed in the administration of these tests. The Aides had to be constantly reminded to follow the standardized methods of administration. Another problem, which did not become apparent until almost the end of the program, was that of keeping the log books. This proved to be too demanding for two of the aides who had only limited writing skills. Unfortunately, they were reluctant to discuss the problem with the Center staff. If they had made the problem known, alternative solutions, such as using a tape recorder, or a more structured check list, might have simplified the task.

The Data Aides proved very competent as Parent Interviewers, having excellent rapport with the Head Start parents. One Data Aide observed that some mothers in a housing project were ill at ease when she arrived to give the interview in formal office attire. She was given permission to wear slacks when making home visits and subsequently found the mothers much more comfortable and more willing to provide the information required in the interview. The Aides reported that often they were identified as social workers and the parents had to be convinced that the interview was being conducted in order to improve the Head Start program. After rapport was established, the mothers were very willing to talk and in fact many of them wanted to talk about aspects of their lives not covered by the Parent Interview.

It was hoped that after the completion of the year's evaluation, these Data Aides would be placed in Head Start positions so that they could take advantage of the skills they had acquired in working with children, teachers, and parents. This hope was only partially fulfilled. One Aide is now an assistant teacher in the Head Start Agency in which she had served as a Data Aide; one stayed on at the E & R Center and is now engaged in another research project with Head Start children; one transferred to the Medical Center at UCLA where she has a position as an intake clerk, helping patients who speak only Spanish; and the fourth aide transferred to a clerical position within the university so as to avoid going through the painful experience of filling out application forms and being subjected to interviews. On the whole the Data Aide program can be considered a success, and should provide a precedent for other agencies doing research in poverty areas.

DATA AIDE

Qualifications:

1. Same ethnic group as majority of students and teachers in class.
2. A member of the community in which the Center is located, if at all possible.
3. High school graduate preferred, but a mature, responsible adult, experienced with her own children, would be considered.

Duties:

For all classes to which assigned, the Data Aide will:

1. Obtain Parent Permission for evaluation;
2. Keep all descriptive data such as child attendance records, teacher attendance records, number and relationship of child visitors, record and identify adult visitors (supervisors, parents, volunteers, special resource people, mental health consultants, social workers, etc.);
3. Keep record of parent meetings, content of meeting and who attended, by name (including children not in Head Start class);
4. Determine when trips are to be made and inform evaluation office of these in advance when such activities may conflict with projected testing or observation sessions;
5. Administer simple test instruments such as Preschool Inventory, Animal House, Gumpgookies, etc.
6. Administer Parent Interview, either at site or in home;
7. Keep inventory of classroom equipment and materials, indicating accessibility to children, as required on Classroom Facilities form.
8. Meet with Coordinator or other UCLA staff member to submit a weekly report and discuss specific problems. This may take place either at the Head Start Site or at UCLA, whichever is most convenient.
9. Maintain good rapport with regular Head Start personnel. The Aide should pitch in if it is clear that there is no attempt to take over the teacher's role, and if this does not interfere with the discharging of the primary responsibility, which is to the UCLA Head Start E & R intervention project.

Results and Discussion

Since the hypotheses of this study are concerned with the effect of feedback on children and teachers, only the relevant measures in the total evaluation package will be analyzed here. For heuristic convenience, the analyses will be divided into three sections: 1) effects on children, 2) effects on teachers, and 3) correlational data.

Before examining the data from the individual child measures, it is important to point out several features of the sample population. Referring to Table 2b on page 26, it is of interest to note that 80% of the children were black and 16% Mexican-American, with only 2% Anglo and 2% Mixed-Other. Only two of the six Anglo children were in the experimental treatment, which had only 11 of the 47 Mexican-American children. The number of boys was greater than that of girls for the sample in general, but the discrepancy was much greater in the experimental treatment where 57% of the population was male and 43% female. With reference to age, the sample was extremely homogeneous, 77% falling within a nine month (48-56) range. Briefly, there were primarily black four-year-old children in this study.

When attempting to relate changes in children to program or treatment variables, it is often assumed that everyone is in the program for an equivalent amount of time. However, the evaluations of the previous years have indicated a tremendous amount of absenteeism and attrition in the California samples. It is quite possible that children could be included in pre and posttesting, yet have attended only sporadically in the interim.

If attendance in Head Start does have an impact, the performance of children who attend infrequently will have a considerably depressant effect on the mean scores. An attempt was made to get at this variable by calculating percent attendance by class. Inspection of the data on Table 3 reveals a wide range, from 65.3% to 91.8%, with equally wide within-class variance as reflected in standard deviations ranging from 2.8 to 11.9. However, there is a confounding of attrition with attendance. Class III A 8 had 90.6% attendance but had lost over half the entering children by the end of the year. Evidently those who remained were the ones who also attended regularly. On the other hand, Class IV A 9 had 90.5% attendance and lost only 16.7% of its enrollees. Unfortunately, data was not available for the entire year for all classes; also attendance could only be related to mean performance on a class basis for the individual child measures.

Both attrition and attendance show a great deal of variation on a class basis; averaged over treatments or delegate agencies these discrepancies tend to diminish, although Agencies I and III had considerably lower attendance and higher attrition than Agencies II and IV, and treatment C had somewhat lower attendance but 10% less attrition. However, contrary to expectation, no consistent relationship can be seen between mean gains on individual measures and either attendance or attrition.

Evidently teachers in the experimental group were no more successful in keeping their children motivated to attend classes. It may be that differential home environment circumstances, which are not examined in this report, may have had an important effect. Much finer techniques are needed, with more elaborate multivariate analyses, before the critical features of these complex interrelationships can be determined.

Analysis of Individual Child Measures

Tables 4-8 present the data on the tests of individual children, by classes; tables 9-12 present the same measures by agencies and treatments. Scores on the vocabulary, verbal, and performance items of the Stanford Binet are given separately by class, treatment, and agency in Tables 4a, b, and c. Although no class shows an actual raw score decrement from pre- to posttest, when the performance over the total test is summed and transformed to an I.Q. score, as shown on Table 5, there are several classes which show slight but non-significant decreases as a result of increased C.A. without commensurate increase in M.A.

A question of interest was whether the children who dropped were measurably different from those who remained for the entire year. Thus on each table the scores of children who were pretested but not posttested are given; in the adjacent columns are the scores of those who had both pre- and posttests. Some discrepancy in the number of subjects given tests is due to the fact that the pre-post population includes only those children who received both tests; several children who were posttested had been either untestable or unavailable for pretesting. Pretest scores of those who dropped out tend to show a wider range of performance compared to those who remained; both the lowest and the highest score on every measure was made by children who dropped.

Table 5 presents the data for the total Binet, as well as the PPVT which was given to the second control group. The scores on the latter test, which measures only verbal skills, are, on the pretest, much lower than the Binet I.Q. scores. However, the gains made on this test tend to be on the average greater than those on the Binet so that the posttest PPVT I.Q.'s are more closely analogous to the Binet I.Q.'s.

The analysis of the Caldwell Preschool Inventory, as shown in Tables 6 and 10, is in terms of raw scores only. The norms provided by the University of Hawaii were not used because of their irregularities; for example, the mean score at 54 months is lower than the mean for either 51 or 53 months. Since approximately 80% of the population was within the 48-54 month age range, it seemed more appropriate to use the raw scores. While all three treatments showed significant pre-post gains on the Caldwell, the experimental treatment had the lowest mean score on the posttest.

On the Animal House subtest of the WPPSI, only three of the 24 classes showed a decrement on the posttest. Two of these classes were above the mean scaled score of 10 on the pretest, with one class having a mean scaled score of 14.5. The posttest score for this class was 11.5, a

Table 4a

Pre, Post, and Gain Scores on Binet Vocabulary Subtest

(With Means for Subjects Dropped before Posttesting)

Agency	Treatment	Classes	Pretest Only			Completed Year's Program						
						Pre			Post		Gain	
			N	M	SD	N	M	SD	M	SD	M	SD
I	A	1	2	3.0	0.0	7	2.7	0.8	3.6	0.5	0.9	0.7
I	A	2	5	3.0	0.7	4	3.3	1.0	3.8	0.5	0.5	1.0
I	A	3	4	2.8	0.4	6	3.3	0.5	3.5	0.5	0.2	0.4
I	A	4	5	2.8	0.4	7	2.6	0.8	3.2	0.8	0.7	0.8
II	A	5	-	-	-	10	2.8	0.8	3.3	0.7	0.5	0.7
II	A	6	1	2.0	0.0	9	3.2	1.0	3.7	1.0	0.4	0.5
III	A	7	2	3.0	0.0	8	3.1	0.4	4.8	3.8	1.6	3.8
III	A	8	5	2.8	1.1	5	3.0	0.7	3.6	0.5	0.6	0.5
IV	A	9	1	3.0	0.0	11	3.1	0.8	3.7	0.5	0.6	0.7
IV	A	10	4	2.3	1.0	6	3.3	0.5	3.7	0.5	0.3	0.5
I	B	11	2	3.5	0.7	8	3.1	0.6	4.9	4.1	2.0	3.9
I	B	12	2	3.0	0.0	7	3.8	0.5	3.9	0.4	0.3	0.5
II	B	13	3	3.0	0.0	8	3.4	0.5	3.7	0.5	0.4	0.7
II	B	14	3	3.7	0.6	4	3.0	0.8	3.3	1.0	0.3	1.0
III	B	15	5	3.2	0.4	6	3.2	0.4	3.5	0.5	0.3	0.8
IV	B	16	3	3.3	0.6	6	3.0	0.0	3.5	0.5	0.5	0.5
IV	B	17	1	4.0	0.0	7	3.1	1.1	3.6	0.5	0.6	0.8
IV	B	18	2	3.0	0.0	9	3.0	0.9	3.7	0.7	0.7	0.5
I	A		16	2.9	0.5	24	2.9	0.8	3.5	0.6	0.6	0.7
II	A		1	2.0	0.0	19	3.0	0.9	3.5	0.8	0.5	0.6
III	A		7	2.8	1.0	13	3.1	0.5	4.3	3.0	1.2	3.0
IV	A		5	2.4	0.9	17	3.2	0.7	3.7	0.5	0.5	0.6
I	B		4	3.3	0.5	15	3.3	0.6	4.4	2.8	1.1	2.7
II	B		6	3.3	0.5	12	3.3	0.6	3.5	0.7	0.3	0.8
III	B		5	3.2	0.4	6	3.2	0.4	3.5	0.5	0.3	0.8
IV	B		6	3.3	0.5	22	3.0	0.8	3.6	0.6	0.6	0.6
I			20	2.9	0.5	39	3.1	0.7	3.8	1.8	0.8	1.7
II			7	3.1	0.7	31	3.1	0.8	3.5	0.8	0.4	0.7
III			12	3.0	0.8	19	3.1	0.5	4.0	2.5	0.9	2.5
IV			11	2.9	0.8	39	3.1	0.8	3.7	0.5	0.6	0.6
	A		29	2.7	0.7	73	3.0	0.7	3.7	1.4	0.7	1.4
	B		21	3.3	0.5	55	3.2	0.7	3.8	1.5	0.6	1.5

Table 4b

Pre, Post, and Gain Scores on Binet Verbal Subtest

(With Means for Subjects Dropped before Posttesting)

Agency	Treatment	Classes	Pretest Only			Completed Year's Program						
						Pre			Post		Gain	
			N	M	SD	N	M	SD	M	SD	M	SD
I	A	1	2	6.5	0.7	7	6.6	1.3	10.0	2.0	3.4	1.6
I	A	2	5	7.0	2.7	4	8.5	1.9	10.5	2.4	2.0	0.8
I	A	3	4	4.5	1.3	6	9.2	3.3	11.3	2.0	2.2	2.2
I	A	4	5	6.0	1.6	7	6.9	3.8	11.3	2.4	4.4	2.7
II	A	5	-	-	-	10	7.8	2.8	9.7	3.5	1.9	2.4
II	A	6	1	6.0	0.0	9	7.0	2.7	10.4	3.4	3.4	2.1
III	A	7	2	13.0	0.0	8	9.3	1.9	10.9	2.7	1.6	1.6
III	A	8	5	8.0	3.2	5	7.4	2.2	11.6	1.8	4.2	0.8
IV	A	9	1	9.0	0.0	11	7.5	2.1	11.4	2.3	3.9	2.2
IV	A	10	4	6.3	2.9	6	10.0	3.6	13.8	3.7	3.8	1.7
I	B	11	2	8.0	0.0	8	9.5	2.9	11.7	2.6	2.3	1.2
I	B	12	2	9.5	0.7	7	11.9	3.2	12.7	3.8	0.9	1.1
II	B	13	3	5.7	1.2	8	9.0	2.4	13.3	2.1	4.8	2.5
II	B	14	3	10.3	4.7	4	8.5	3.0	12.3	2.5	3.8	1.3
III	B	15	5	8.8	3.0	6	7.7	1.4	12.8	2.6	5.2	2.9
IV	B	16	3	8.7	3.8	6	8.2	2.7	12.7	1.8	4.5	2.7
IV	B	17	1	10.0	0.0	7	8.6	2.2	11.9	2.0	3.6	1.3
IV	B	18	2	8.5	2.1	9	8.4	3.6	11.1	2.8	2.7	1.7
I	A		16	6.0	2.0	24	7.6	2.9	10.8	2.1	3.2	2.9
II	A		1	6.0	0.0	19	7.4	2.7	10.0	3.4	2.6	2.3
III	A		7	9.4	3.6	13	8.5	2.1	11.2	2.3	2.6	1.9
IV	A		5	6.8	2.8	17	8.4	2.9	12.2	3.0	3.9	2.0
I	B		4	8.8	1.0	15	10.6	3.2	12.2	3.2	1.5	1.3
II	B		6	8.0	4.0	12	8.8	2.5	13.0	2.2	4.4	2.2
III	B		5	8.8	3.0	6	7.7	1.4	12.8	2.6	5.2	2.9
IV	B		6	8.8	2.6	22	8.4	2.8	11.8	2.3	3.5	2.0
I			20	6.6	2.1	39	8.8	3.3	11.3	2.6	2.6	2.1
II			7	7.7	3.7	31	8.0	2.7	11.3	3.3	3.3	2.4
III			12	9.2	3.2	19	8.3	1.9	11.7	2.5	3.4	2.5
IV			11	7.9	2.8	39	8.4	2.8	12.0	2.6	3.6	2.0
	A		29	7.0	2.8	73	7.9	2.7	11.0	2.8	3.1	2.1
	B		21	8.6	2.8	55	9.0	2.9	12.3	2.4	3.4	2.3

Table 4c

Pre, Post, and Gain Scores on Binet Performance Subtest

(With Means for Subjects Dropped before Posttesting)

Agency	Treatment	Classes	Pretest Only			Completed Year's Program						
						Pre			Post		Gain	
			N	M	SD	N	M	SD	M	SD	M	SD
I	A	1	2	19.5	2.1	7	15.3	2.7	21.4	1.3	6.1	2.0
I	A	2	5	17.2	5.3	4	18.5	3.0	22.0	3.5	3.5	1.9
I	A	3	4	12.0	2.0	6	18.8	4.5	23.2	3.2	4.3	3.3
I	A	4	5	16.2	2.8	7	15.9	3.8	23.3	2.4	7.4	2.9
II	A	5	-	-	-	10	15.6	3.2	20.9	3.8	5.3	1.3
II	A	6	1	18.0	0.0	9	16.9	3.6	21.0	1.9	4.1	2.2
III	A	7	2	24.5	1.5	8	17.8	2.6	21.6	3.3	3.9	4.9
III	A	8	5	17.8	6.7	5	19.0	3.8	22.2	3.2	3.2	1.1
IV	A	9	1	20.0	0.0	11	18.3	2.6	22.4	2.2	4.1	1.9
IV	A	10	4	15.5	4.0	6	21.2	3.4	20.3	4.8	-0.8	6.2
I	B	11	2	21.0	0.0	8	19.0	3.4	22.9	1.9	2.8	3.3
I	B	12	2	20.0	1.4	7	21.3	2.4	23.9	3.0	2.6	1.9
II	B	13	3	16.0	1.0	8	19.3	3.1	22.9	2.1	3.9	3.1
II	B	14	3	19.0	4.4	4	16.5	5.8	23.3	2.5	6.8	4.3
III	B	15	5	19.2	2.2	6	17.5	2.7	23.0	0.9	5.5	3.1
IV	B	16	3	17.3	4.2	6	19.8	1.0	24.2	1.5	4.3	0.8
IV	B	17	1	20.0	0.0	7	18.6	2.4	23.1	1.6	4.6	2.1
IV	B	18	2	20.0	4.2	9	18.2	2.5	23.1	1.7	4.9	3.1
I	A		16	16.3	4.0	24	16.9	3.7	22.5	2.5	5.6	2.9
II	A		1	18.0	0.0	19	16.2	3.4	20.9	2.9	4.7	1.9
III	A		7	18.7	6.4	13	18.2	3.0	21.8	3.2	3.6	3.8
IV	A		5	16.4	4.0	17	19.3	3.2	21.6	3.3	2.4	4.5
I	B		4	20.5	1.0	15	20.1	3.1	23.4	2.5	2.7	2.5
II	B		6	17.5	3.3	12	18.3	4.1	23.0	2.1	4.8	3.6
III	B		5	19.2	2.2	6	17.5	2.6	23.0	0.9	5.5	3.1
IV	B		6	18.7	3.6	22	18.8	3.1	23.4	1.6	4.6	2.2
I			20	17.2	4.0	39	18.1	3.8	22.8	2.5	4.6	3.1
II			7	17.6	3.0	31	17.0	3.8	21.8	2.8	4.8	2.6
III			12	18.9	4.7	19	18.0	2.9	22.2	2.7	4.2	3.6
IV			11	17.6	3.8	39	19.0	3.1	22.7	2.6	3.6	3.6
	A		29	16.9	4.5	73	17.5	3.5	21.8	2.9	4.3	3.5
	B		21	18.8	2.8	55	18.9	3.3	23.3	1.9	4.3	2.9

Table 5

Means, Standard Deviations, Gain Scores, and Significance Of Difference on Pre-Post Stanford Binet[a] by Classes

			Pretest Only			Completed Year's Program						
						Pre			Post		Gain	
Agency	Treatment	Classes	N	M	SD	N	M	SD	M	SD	M	SD
I	A	1	2	89.5	0.7	7	82.9	7.4	89.0	5.8	6.1	6.4
I	A	2	5	86.0	9.6	4	91.3	12.0	86.8	12.7	-4.5	1.0
I	A	3	5	73.5	5.9	6	90.0	14.4	96.8	9.9	6.8	19.9
I	A	4	5	83.6	12.9	7	77.7	11.6	87.7	7.6	10.0	6.9
II	A	5	-	-	-	10	81.4	12.2	83.5	15.3	2.1	6.5
II	A	6	1	85.0	0.0	9	88.1	10.3	90.8	7.1	2.7	5.4
III	A	7	2	106.0	12.7	8	91.6	8.8	93.0	12.7	1.8	11.6
III	A	8	6	85.8	18.1	5	90.6	12.4	95.2	11.2	4.6	5.1
IV	A	9	1	88.0	0.0	11	88.3	9.4	90.1	10.2	1.8	7.5
IV	A	10	4	81.3	16.2	6	96.5	10.7	103.3	27.9	6.8	17.4
I	B	11	2	93.5	2.1	6	95.5	11.3	92.5	7.5	-3.0	6.0
I	B	12	2	96.0	0.0	7	103.7	7.4	101.6	11.3	-2.1	11.2
II	B	13	3	83.3	6.4	8	92.3	11.2	100.1	8.6	7.9	11.8
II	B	14	3	94.0	18.1	4	88.6	10.3	98.8	5.2	10.0	7.5
III	B	15	5	88.4	13.0	6	84.3	3.2	97.0	9.6	12.7**	8.5
IV	B	16	3	83.7	12.2	6	88.5	8.0	94.2	8.3	5.7	6.6
IV	B	17	1	105.0	0.0	7	90.1	8.2	92.0	4.7	1.9	6.0
IV	B	18	2	89.5	6.4	9	87.4	14.5	89.3	8.4	1.9	8.0
II	C	19	-	-	-	7	57.3	14.7	76.0	13.7	18.7*	9.6
II	C	20	-	-	-	8	83.6	14.7	83.6	16.9	0.0	11.9
III	C	21	-	-	-	6	65.5	12.5	76.7	23.3	11.2	12.2
III	C	22	-	-	-	7	84.7	16.8	90.7	25.5	6.0	21.0
IV	C	23	-	-	-	6	80.0	16.4	104.2	10.1	24.2**	13.6
IV	C	24	-	-	-	9	83.6	16.4	99.3	26.0	15.8	21.1

[a] Classes 19-24 received PPVT instead of Binet.

*$P < .05$
**$P < .01$

Table 6

Means, Standard Deviations, Gain Scores and Significance Of Difference on Pre-Post Caldwell by Classes

Agency	Treatment	Classes	Pretest Only			Completed Year's Program						
						Pre			Post		Gain	
			N	M	SD	N	M	SD	M	SD	M	SD
I	A	1	1	32.0	0.0	7	30.1	6.7	39.6	5.8	9.4*	3.0
I	A	2	4	35.3	10.2	3	35.0	1.0	49.3	4.5	14.3**	5.5
I	A	3	5	26.4	8.0	4	35.8	11.4	47.3	5.5	11.5	7.2
I	A	4	5	34.2	10.2	6	30.8	9.8	38.6	11.2	7.8	3.7
II	A	5	-	-	-	9	31.4	6.3	39.1	9.7	7.7	9.4
II	A	6	-	-	-	9	34.9	5.3	44.8	8.8	9.9**	6.5
III	A	7	2	44.5	6.4	8	32.9	6.7	38.9	8.6	6.0	11.0
III	A	8	6	29.5	11.8	4	42.3	10.1	43.3	4.3	1.0	7.4
IV	A	9	-	-	-	11	31.0	11.8	39.3	8.3	8.2	10.8
IV	A	10	4	23.0	5.3	6	37.3	18.0	44.8	13.1	7.5	6.9
I	B	11	2	32.0	2.8	8	36.8	15.0	49.3	9.4	12.5	8.7
I	B	12	2	32.5	2.1	6	34.6	15.3	50.7	8.4	16.0*	9.7
II	B	13	2	14.5	2.1	9	32.6	14.3	43.8	7.5	11.2	10.7
II	B	14	2	39.0	15.6	4	40.8	15.4	45.8	9.9	5.0	6.5
III	B	15	4	28.8	3.2	5	30.2	11.8	42.8	10.8	12.6	8.8
IV	B	16	2	24.5	20.5	6	34.7	8.9	44.0	6.3	9.3	4.4
IV	B	17	-	-	-	8	26.4	9.4	36.9	12.2	10.5	7.6
IV	B	18	1	47.0	0.0	9	30.9	9.7	39.4	6.7	8.6*	6.1
II	C	19	-	-	-	9	28.7	11.4	43.1	8.8	14.4**	6.8
II	C	20	-	-	-	7	29.3	9.6	42.1	10.8	12.9	10.0
III	C	21	-	-	-	5	30.2	12.6	37.2	6.0	7.0	12.3
III	C	22	-	-	-	8	31.9	10.1	38.5	7.1	6.6	10.2
IV	C	23	-	-	-	6	36.8	9.0	45.5	6.4	8.7	6.5
IV	C	24	-	-	-	10	39.1	12.3	43.2	8.9	4.1	7.1

*P< .05
**P< .01

Table 7

Means, Standard Deviations, Gain Scores, and Significance
Of Difference on Pre-Post WPPSI by Classes

Agency	Treatment	Classes	Pretest Only			Completed Year's Program						
						Pre			Post		Gain	
			N	M	SD	N	M	SD	M	SD	M	SD
I	A	1	1	11.0	0.0	7	10.1	1.7	11.6	3.5	1.4	2.8
I	A	2	4	11.5	3.1	3	12.0	2.6	12.0	6.1	0.0	7.5
I	A	3	5	8.6	4.7	4	14.5	3.1	11.5	4.4	-3.0	4.2
I	A	4	5	11.6	2.7	6	11.8	3.4	11.0	2.4	-0.8	2.6
II	A	5	-	-	-	10	7.1	1.7	9.6	4.0	2.5	3.8
II	A	6	-	-	-	9	8.3	2.3	11.0	2.1	2.7*	2.6
III	A	7	2	9.0	1.4	8	7.1	1.8	10.4	2.5	3.3**	2.9
III	A	8	6	7.8	4.4	5	11.0	3.6	13.0	6.0	2.0	5.5
IV	A	9	-	-	-	11	8.7	3.5	10.1	2.7	1.4	3.2
IV	A	10	4	8.0	2.2	6	11.8	3.1	13.0	3.3	1.2	2.8
I	B	11	2	10.0	5.7	8	11.6	4.1	12.6	2.7	1.0	2.6
I	B	12	2	8.0	1.4	6	11.6	5.2	13.0	2.6	1.3	4.5
II	B	13	2	7.5	0.7	9	8.9	3.0	11.0	3.0	2.1	3.1
II	B	14	2	5.0	4.2	4	9.5	2.4	9.0	3.4	-0.5	3.0
III	B	15	4	9.0	6.2	5	12.0	1.4	13.2	3.3	1.2	3.4
IV	B	16	2	8.0	1.4	6	12.2	5.2	14.5	2.9	2.3	3.3
IV	B	17	-	-	-	8	9.4	2.3	12.4	2.4	3.0*	1.7
IV	B	18	1	12.0	0.0	9	10.6	2.4	13.1	3.1	2.6	3.2
II	C	19	-	-	-	8	9.4	3.2	14.1	2.3	4.8**	3.4
II	C	20	-	-	-	6	7.5	1.5	11.7	4.8	4.2	5.5
III	C	21	-	-	-	5	9.4	3.3	11.0	2.2	1.6	4.0
III	C	22	-	-	-	8	10.3	3.5	11.3	3.1	1.0	2.6
IV	C	23	-	-	-	6	10.5	3.5	13.7	3.3	3.2	2.1
IV	C	24	-	-	-	8	9.5	3.1	12.4	4.0	2.9	4.3

*P<.05
**P<.01

Table 8

Means Standard Deviations, Gain Scores, and Significance Of Difference on Pre-Post Gumpgookies by Classes

Agency	Treatment	Classes	Pretest Only			Completed Year's Program						
						Pre			Post		Gain	
			N	M	SD	N	M	SD	M	SD	M	SD
I	A	1	1	32.0	0.0	5	42.6	4.1	40.4	4.6	-2.2	7.9
I	A	2	3	34.0	2.6	3	33.7	5.9	44.3	2.9	10.7*	3.1
I	A	3	5	37.6	5.4	6	38.6	4.3	47.8	4.0	9.2**	4.6
I	A	4	5	37.8	4.8	7	32.6	8.9	42.9	5.7	10.3*	8.1
II	A	5	-	-	-	10	33.3	4.4	44.8	6.1	11.5**	6.9
II	A	6	-	-	-	9	34.1	11.3	44.3	6.4	10.2	15.2
III	A	7	2	38.5	4.9	8	33.9	6.6	38.9	6.6	5.0	8.4
III	A	8	2	33.5	10.6	4	33.5	4.0	48.3	3.8	14.8**	4.6
IV	A	9	-	-	-	11	28.0	7.7	37.8	8.8	9.8*	11.7
IV	A	10	3	32.3	1.5	6	30.2	15.3	39.3	7.3	9.2	12.3
I	B	11	2	37.0	1.4	7	35.3	5.8	43.0	7.2	7.7*	2.9
I	B	12	2	42.0	2.8	6	38.8	4.6	47.0	5.3	8.2*	2.5
II	B	13	1	27.0	0.0	7	36.6	2.7	40.4	6.2	0.5	10.0
II	B	14	2	38.5	2.1	4	32.8	8.0	43.3	5.7	10.5	7.5
III	B	15	4	38.8	4.7	6	33.3	9.2	42.0	9.4	8.7	13.4
IV	B	16	2	28.5	4.9	6	36.0	7.7	42.5	7.0	6.5	5.7
IV	B	17	-	-	-	7	30.3	4.5	39.4	8.1	9.1*	8.1
IV	B	18	1	36.0	0.0	9	29.7	4.9	41.9	7.8	12.2**	6.0
II	C	19	-	-	-	7	39.4	3.5	41.7	6.8	2.3	8.5
II	C	20	-	-	-	6	33.8	4.0	36.2	5.0	2.3	2.6
III	C	21	-	-	-	5	31.6	4.3	40.6	3.4	9.0**	4.5
III	C	22	-	-	-	5	36.4	5.0	41.6	7.8	5.2	9.2
IV	C	23	-	-	-	6	30.5	3.9	40.0	8.8	9.5*	7.8
IV	C	24	-	-	-	8	34.4	5.2	40.6	5.9	6.3*	7.3

*P$<$.05
**P$<$.01

Table 9

Means, Standard Deviations, Gain Scores, and Significance of Difference on Pre-Post Stanford Binet[a] By Sites, Agencies, and Treatments

Agency	Treatment	Number of Classes	Pretest Only			Completed Year's Program						
						Pre			Post		Gain	
			N	M	SD	N	M	SD	M	SD	M	SD
I	A	4	17	80.6	10.4	24	84.5	12.0	90.2	9.1	5.7	11.6
II	A	2	1	85.0	0.0	19	84.6	11.5	86.9	12.4	2.4	5.8
III	A	2	8	91.6	18.5	13	91.2	9.9	93.8	11.7	2.6	9.5
IV	A	2	5	82.6	14.3	17	91.2	10.4	94.8	18.7	3.6	11.7
I	B	2	4	94.8	1.9	15	98.1	11.4	97.3	10.1	-2.5	8.8
II	B	2	6	88.7	13.5	13	91.1	10.5	99.3	7.3	8.6	10.3
III	B	1	5	88.4	13.0	6	84.3	3.2	97.0	9.6	12.7	8.5
IV	B	3	6	89.2	11.7	23	88.6	10.8	91.0	7.5	2.9	6.9
I		6	21	85.0	10.6	40	89.8	13.4	92.8	9.9	2.8	11.3
II		4[b]	7	88.1	12.4	32	87.1	11.4	92.0	12.1	4.8	8.3
III		3[b]	13	90.3	13.9	19	89.1	8.9	94.8	10.9	5.8	10.1
IV		5[b]	11	86.2	12.7	40	89.7	10.5	92.6	13.4	3.2	9.2
	A	10	31	84.8	13.3	73	87.3	11.4	91.1	13.2	3.8	9.9
	B	8	21	89.9	10.9	53	91.5	11.0	95.4	8.9	3.9*	9.6
	C	6	-	-	-	43	76.5	17.9	88.7	22.0	12.2**	-

[a] Group C received PPVT instead of Binet
[b] Does not include C-2 classes

*p <.05
**p <.01

Table 10

Means, Standard Deviations, Gain Scores, and Significance of Difference on Pre-Post Caldwell By Sites, Agencies, and Treatments

Agency	Treatment	Number of Classes	Pretest Only			Completed Year's Program						
						Pre			Post		Gain	
			N	M	SD	N	M	SD	M	SD	M	SD
I	A	4	15	31.7	9.3	23	32.8	8.5	41.7	8.7	10.1	4.8
II	A	2	-	-	-	18	33.1	5.9	41.7	9.2	8.8	7.9
III	A	2	8	33.3	12.4	12	36.0	8.8	38.8	9.2	4.3	9.9
IV	A	2	5	24.6	5.8	17	33.2	14.0	41.2	10.2	8.0	9.4
I	B	2	4	32.3	2.1	16	35.1	13.7	49.9	8.7	14.0	8.9
II	B	2	4	26.8	16.8	13	35.1	14.5	44.4	7.9	9.3	9.8
III	B	1	4	28.8	3.2	6	28.8	11.1	42.8	10.8	12.6	8.8
IV	B	3	4	30.3	16.3	23	30.3	9.5	39.7	9.0	9.4	6.0
I		6	19	31.8	8.2	39	33.8	10.8	44.9	9.5	11.7	7.0
II		4[a]	5	28.2	15.0	31	34.0	10.2	42.8	8.7	9.0	8.6
III		3[a]	12	31.8	10.3	18	33.6	9.9	39.9	9.5	6.8	10.1
IV		5[a]	9	27.1	11.2	40	31.6	11.6	40.4	9.4	8.8	7.6
	A	10	29	31.0	9.8	70	33.6	9.5	41.0	9.2	8.2**	8.0
	B	8	16	29.5	10.8	55	32.8	12.3	43.7	9.5	10.9**	8.1
	C	6	-	-	-	45	32.9	11.2	41.8	8.3	8.9**	-

[a] Does not include C-2 classes

**$p < .01$

Table 11

Means, Standard Deviations, Gain Scores, and Significance of Difference on Pre-Post WPPSI by Sites, Agencies, and Treatments

Agency	Treatment	Number of Classes	Pretest Only			Completed Year's Program						
						Pre			Post		Gain	
			N	M	SD	N	M	SD	M	SD	M	SD
I	A	4	15	10.5	3.5	23	11.8	2.8	11.2	3.5	-0.4	4.0
II	A	2	-	-	-	19	7.7	2.0	10.3	3.3	2.6	3.2
III	A	2	8	8.1	3.8	13	8.6	3.2	11.4	4.2	2.8	3.9
IV	A	2	5	7.6	2.1	17	9.8	3.6	11.1	3.2	1.3	3.0
I	B	2	4	9.0	3.6	16	11.6	4.3	12.8	2.6	1.1	3.4
II	B	2	4	6.3	2.9	13	9.1	2.7	10.4	3.2	1.3	3.2
III	B	1	4	9.0	6.2	6	12.2	1.3	13.2	3.3	1.2	3.4
IV	B	3	4	8.8	3.4	23	10.6	2.3	13.2	2.8	2.7	2.7
I		6	19	10.2	3.5	39	11.7	3.5	11.9	3.2	0.3	2.8
II		4^{a}	5	6.0	2.5	32	8.3	2.4	10.3	3.1	9.1	3.2
III		3^{a}	12	8.4	4.4	19	9.7	3.1	11.9	3.9	2.3	3.8
IV		5^{a}	9	8.1	2.1	40	10.3	3.4	12.3	3.1	2.1	2.9
	A	10	29	9.2	3.6	68	9.7	3.1	11.0	3.4	1.4*	3.7
	B	8	16	8.3	3.8	55	10.6	3.4	12.4	3.0	1.8**	3.1
	C	6	-	-	-	41	9.5	3.1	12.4	3.4	2.9**	-

[a] Does not include C-2 classes

*p < .05
**p < .01

Table 12

Means, Standard Deviations, Gain Scores, and
Significance of Difference on Pre-Post Gumpgookies
By Sites, Agencies, and Treatments

Agency	Treatment	Number of Classes	Pretest Only			Completed Year's Program						
						Pre			Post		Gain	
			N	M	SD	N	M	SD	M	SD	M	SD
I	A	4	14	36.5	4.6	23	37.3	7.0	44.0	5.1	7.0	8.3
II	A	2	-	-	-	19	33.7	8.2	44.6	6.1	10.9	11.3
III	A	2	4	36.0	7.3	12	33.8	5.7	41.7	7.1	8.3	8.6
IV	A	2	4	33.8	3.1	17	28.8	10.5	38.4	8.1	9.6	11.5
I	B	2	4	39.5	3.4	16	35.6	5.6	44.9	6.5	7.9	2.6
II	B	2	3	34.7	6.8	12	34.8	5.4	38.4	10.8	3.8	10.1
III	B	1	4	38.8	4.7	6	33.3	9.2	42.0	9.4	8.7	13.4
IV	B	3	4	32.3	5.2	22	31.6	6.1	41.0	7.4	9.7	6.8
I		6	18	37.2	4.5	39	36.6	6.5	44.3	5.6	7.3	6.6
II		4[a]	4	37.0	7.3	31	34.1	7.1	42.1	8.7	8.2	11.2
III		3[a]	8	37.4	5.9	18	33.6	6.8	41.8	7.6	8.4	10.1
IV		5[a]	8	33.0	4.0	39	30.4	8.3	39.9	7.7	9.6	9.0
	A	10	23	36.3	5.0	64	35.0	5.8	42.5	6.9	8.9**	10.0
	B	8	15	36.4	5.4	51	34.3	5.9	42.2	7.1	7.8**	8.0
	C	6	-	-	-	37	34.5	5.1	40.1	6.5	5.6**	-

[a] Does not include C-2 classes

**$p < .01$

three-point loss, yet still above average. On the posttest, all except two classes (scaled scores of 9.0 and 9.6) were performing above the mean. However, there was again no relationship between performance and treatment condition.

Of all the individual child measures, the Gumpgookies showed the most classes with significant pre-post gains. However, this may very well be due to the fact that the 55-item posttest was compared to the same items embedded in a 100-item pretest. Fatigue and position effects were probably much more operative on the longer test. As with the Caldwell, raw scores rather than the transformed scores were used, for the same reasons. (E.g., the mean score for 44 months was 34.2 and that for 51 months 33.9.)

While there is wide variation among classes and agencies on these individual child measures, analysis of variance on gains and of covariance on posttest scores (Tables 13 and 14) confirm the failure to find significant differences among treatment groups on any of these measures. It may be that some of the within-treatment variance could have been due to the conditions under which the testing was done. This was known to vary considerably from one site to the next and even between classes at the same site. Thus, as can be seen in Table 15, at one site all four classes were tested in the same trailer, which had been placed on the playground area for the duration of the test period. However, two of the classes were morning classes so that the testing took place when there was comparatively little noise on the playground; in the afternoon the noise level was beyond the control either of the examiner or the Head Start staff. While no statistical tests of the relationship between test conditions and test scores were run, class III B 15, which was slightly below the average on the Binet but had the best testing conditions, scored among the highest on the Caldwell and the WPPSI. Unfortunately, the relationships between scores and testing conditions were too weak to produce any reliable comparisons across treatments. Whatever may have been the contributing reasons, teachers receiving feedback information were not more effective in producing measurably greater progress in the behaviors tested.

In addition to the national core battery for individual children, a series of behavioral items specifically designed to test teacher effectiveness in terms of their most highly valued learnings was administered to a small sample. The rationale and description of this instrument, and the results, may be found in detail in the research paper included on pages 174 to 178 of this Final Report. Briefly, the major findings were that children learn best what their teachers value most--at least at this early age!

Analysis of Teacher and Classroom Measures

Just as attendance and attrition of the children have often been disregarded in assessing the impact of interventions via pre-post designs, so have these factors been overlooked in determining the differential

Table 13

One-Way Analyses of Variance on Gain Scores for Individual Measures (over 24 classes)

Dependent Variable	Source of Variance	df	Mean Square	F
Caldwell	Treatment	23	68.23	.97
	Error	143	70.11	
Gumpgookies	Treatment	23	66.26	1.22
	Error	128	54.39	
WPPSI	Treatment	23	16.11	1.35
	Error	140	11.91	
I.Q.[a]	Treatment	23	323.65	2.50
	Error	145	129.66	

Table 14

Pre-Post Analyses of Covariance for Individual Measures (over 24 classes)

Dependent Variable	Source of Variance	df	Mean Square	F
Caldwell	Treatment	23	62.05	1.45
	Error	142	42.71	
Gumpgookies	Treatment	23	48.00	1.12
	Error	127	42.73	
WPPSI	Treatment	23	9.35	1.01
(Animal House)	Error	139	9.22	
I.Q.[a]	Treatment	23	206.78	1.68
	Error	144	123.22	
Binet only	Treatment	17	100.52	1.29
	Error	107	77.96	

---

[a] Based on Binet for Experimental and Control 1; PPVT for Control 2.

Table 15

Testing Conditions

Agency	Treatment	Classes	Type of Facility	Location	Accessibility to Classroom	Freedom from Noise	Working Space	Working Surface	Lighting	Ventilation	Temperature	Cleanliness	Overall Rating
I	A	1	a	f	3	1	2	2	2	2	2	2	2
I	A	2	a	f	3	2	2	2	2	2	2	2	2
I	A	3	a	f	3	1	2	2	2	2	2	2	1
I	A	4	a	f	3	2	2	2	2	2	2	2	2
II	A	5	c	h	2	1	1	1	2	2	1	2	2
II	A	6	c	h	2	2	2	1	2	2	1	2	2
III	A	7	b	g	3	2	2	2	3	3	3	3	3
III	A	8	b	g	2	2	3	3	3	3	3	3	3
IV	A	9	d	i	2	2	2	2	2	2	2	2	2
IV	A	10	d	i	2	2	2	2	2	2	2	2	2
I	B	11	b	g	3	2	2	2	2	2	2	2	3
I	B	12	b	g	3	2	2	2	2	3	3	2	2
II	B	13	e	f	3	2	3	2	3	3	2	3	3
II	B	14	e	f	3	2	3	3	3	2	2	3	3
III	B	15	b	g	3	3	3	3	3	3	3	3	4
IV	B	16	c	i	2	3	1	1	2	2	1	3	1
IV	B	17	d	i	3	2	2	1	2	2	2	3	2
IV	B	18	c	i	2	1	2	2	2	1	1	2	2

Descriptions:

a = Trailer
b = Room
c = Station Wagon
d = Experimenter's Car
e = Outdoor Table
f = Playground
g = Center Building
h = Parking Lot
i = Street

Rating Scale:

1 = Poor; Data likely to be impaired
2 = Fair but acceptable
3 = Good
4 = Optimal

The Overall Rating is a consensus of subjective estimates of the general atmosphere during testing and is not an average over the individual items.

effectiveness of the classroom teacher. In California, not only are the families of Head Start children extremely mobile, but also the turnover of teaching staff at all grade levels is higher than in most parts of the United States. It seems that this instability on the part of the teacher should have an impact on the gains made by the children in that class, irrespective of the innovative program being introduced. Where the entire focus of the intervention is the classroom teacher, the effect of teacher stability should have an even greater impact.

In the present study, it was possible to determine that 15 of 20 Head Teachers and 15 of 20 Assistant Teachers who attended the Orientation Meeting in October also attended the final meeting in June. Of the five teachers who were replaced, three were in the experimental, one from the primary control, and one from the second control group. While these data alone tell little about the quality and regularity of teacher attendance, at least there seems to have been more continuity than had been expected. The relatively small attrition in all groups may account for the fact that no consistent relationship between teacher turnover and achievement of children was found.

For the experimental group, records of attendance at the feedback meetings were available. These are summarized and presented in Table 16, under the following five categories: Head Teacher, Aide, Child Development Supervisor, Delegate Agency Representative, and UCLA Head Start Center representative. The statistics presented are based on the number of people attending per class for teachers and aides, and per meeting for the other categories. Several sites had more classes than were selected for the feedback study. At one site where there were two sample classes, the teachers and aides from all five classes attended the orientation meeting. However, the pattern of attendance at the subsequent feedback sessions is a familiar one: initial enthusiasm followed by a gradual falling off of interest, with a slight upward swing for the last meeting.

It should be noted that only with Delegate Agency II was there a combined final summary meeting of all the treatment groups. Because of the difficulty of setting a time at which all the teachers could attend, and the problem of presenting each teacher with meaningful feedback about the children in her own class, at all other sites the last meeting was on an individual basis. The Evaluation Coordinator visited the class, provided appropriate feedback, and administered the posttests with the TEACH and TATE instruments.

Characteristics of Teaching Staff

Although this instrument was administered in October and in May to the Head Teacher of each class, the data reported here are from the posttest only. The explanation of the codes used on Table 17 are presented on page 81. It will be noted that on this and subsequent tables data is reported for only 18 classes (Groups A and B). To keep evaluation contacts to a minimum in Group C, no observations were carried out in those classes.

Table 16

Number of People Attending each Meeting

Agency	Treatment	No. of Classes	No. of Meetings	Attendance				
				Per Class		Per Meeting		
				Head Teacher	Aide	UCLA	Child Development Supervisor	Delegate Agency Representative
I	A	4	8	0.91	0.72	2.12	0.38	0.25
II	A	2	8	1.00	0.94	3.12	0.88	0.63
III	A	2	8	1.13	0.81	2.38	0.38	0.25
IV	A	2	8	0.88	1.06	2.25	0.00	0.50
I	B	2	1	1.00	0.50	4.00	0.00	0.00
II	B	2	3[a]	1.00	1.00	2.00	0.67	0.00
III	B	1	1	1.00	1.00	3.00	1.00	0.00
IV	B	3	1	1.00	0.33	3.00	0.00	0.00
II	C	2	2	1.00	1.00	2.00	0.00	0.00
III	C	2	1	1.00	1.00	2.00	0.00	0.00
IV	C	2	1	1.00	0.50	3.00	0.00	0.00

---

[a] The Head Teacher of a Control class attended the Feedback Meeting at UCLA as a substitute Child Development Supervisor!

Explanation of Description and Coding for Items On the "Characteristics of Teaching Staff" and New Composite Variables

Item on Protocol

3. Age
   1 = Under 16
   2 = 16 - 21
   3 = 22 - 27
   4 = 28 - 33
   5 = 34 - 39
   6 = 40 - 45
   7 = 46 - 51
   8 = 52 - 57
   9 = Over 57

4. Sex
   1 = Male
   2 = Female

5. Ethnicity
   1 = Negro
   2 = Mexican Amer.
   3 = Puerto Rican
   4 = White
   5 = Amer. Indian
   6 = Oriental
   7 = Eskimo
   8 = Polynesian
   9 = Mixed, Other

6. First Language; 7. Language Used in Class
   1 = Standard English
   2 = Mexican-Spanish
   3 = Puerto Rican or Cuban Spanish
   4 = Oriental
   5 = Amer. Indian
   6 = Polynesian
   7 = Dialect and/or Non-Standard Engl.
   8 = Other Language
   9 = Standard + Other (Bilingual)

8. General Education
   0 = Grade school diploma
   1 = Some high school
   2 = High school graduate
   3 = Some college
   4 = A.A. degree
   5 = B.A. or B.S.
   6 = Course credit beyond B.A.
   7 = M.A.
   8 = Credit above M.A.
   9 = Doctorate

9. Early Childhood Education
   1 = None
   2 = A.A.
   3 = Undergrad. courses
   4 = B.A. or B.S.
   5 = Graduate courses
   6 = M.A. or M.S.
   7 = Doctorate

10. Head Start Training
   1 = None
   2 = Less than 2 weeks
   3 = 2-3 weeks
   4 = 4-5 weeks
   5 = 6-7 weeks
   6 = 8 weeks
   7 = 8 weeks + seminar
   8 = Other

11. Special training program (most recent)
   1 = None
   2 = Montessori
   3 = Bank Street
   4 = Bereiter
   5 = Nimnicht
   6 = Gray-Miller
   7 = Sprigle
   8 = Other (Specify)

12. No. of months of special training under Item 11
   1 = None
   2 = 1 or less
   3 = 2
   4 = 3 - 6
   5 = 7 - 9
   6 = 10 - 11
   7 = 12 - 18
   8 = 19 - 24
   9 = Over 24

New Composite Variables Used in UCLA Analysis. Months of:
   a. Experience with children other than Head Start. Items 13 - 16.
   b. Head Start only. Items 17 & 18.
   c. All children older than Head Start. Items 13 & 15.
   d. All preschool children, exclusive of Head Start. Items 14 & 16.
   e. All disadvantaged preschool children. Items 16 - 18.

Table 17

Characteristics of Teaching Staff

Item No.	Ratings by Classes																		Experimental		Control	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	M	SD	M	SD
3.	4	8	5	5	5	5	5	9	5	6	4	3	5	7	5	8	4	3	5.7	1.4	4.9	1.6
4.	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	-	-	-	-
5.	1	1	1	1	1	4	4	1	1	1	1	4	1	1	1	2	1	2	-	-	-	-
6.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	-	-	-	-
7.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	9	9	9	-	-	-	-
8.	3	3	3	5	3	5	3	5	3	4	6	6	5	3	5	1	3	3	3.7	0.9	4.1	1.6
9.	3	3	3	3	3	3	3	4	2	1	5	1	5	3	4	1	1	2	2.8	0.7	2.8	1.7
10.	1	3	1	8	8	7	8	7	6	1	8	1	7	8	1	6	6	6	5.4	2.9	5.5	2.5
11.	8	3	1	1	3	8	1	1	8	4	8	8	3	8	8	1	2	3	4.2	3.1	4.9	2.9
12.	5	3	1	1	3	2	1	1	5	2	4	4	9	9	6	1	4	4	2.6	1.6	5.0	2.5
Composite Variables (In Months)																						
a.	70	0	24	48	48	102	178	70	0	24	27	27	18	280	79	24	0	0	56.4	50.8	56.8	87.5
b.	34	60	58	58	34	34	34	34	34	34	34	8	34	34	64	64	34	34	41.4	11.3	38.3	17.0
c.	0	0	24	24	48	48	54	70	0	0	24	24	9	140	9	0	0	0	26.8	25.3	25.8	44.1
d.	78	0	0	24	0	54	124	0	0	24	0	3	9	140	70	24	0	0	29.6	39.5	30.8	46.9
e.	104	60	58	58	34	88	88	34	34	34	34	11	43	104	64	88	34	34	50.4	28.5	52.6	28.8

The mean age of teachers in the experimental group is approximately 36 years, whereas that of the control group is 31 years. There was one teacher over 57 and two between 52 and 57 years of age; no teacher was less than 22.

All but one of the teachers were female, 13 of the 18 were black, two Mexican-American, and four white. Of the latter group, two were in the experimental treatment. All the teachers in the experimental group spoke English as a first language; two of the Mexican-American teachers spoke Spanish as a first language and three of them were bilingual; one of the bilingual teachers was black. Except for one teacher who had only had some high school, all the teachers were at least high school graduates and some had had college training. One had an A.A. degree and five had B.A. degrees.

To try to get at differences in experience of teachers, the actual months for items which seemed to be related were taken as the mid-point of each category and summed for each teacher. Table 17 shows that one teacher in the control group had over 23 years of experience with children other than Head Start, one teacher in the experimental group had almost 15 years of such experience, and two teachers in each group had none. Five of the 18 teachers reported between 58 and 64 months of Head Start experience. Inasmuch as Head Start had been in operation only 48 months as of June, 1969, and for six of these months there was no Head Start in the Los Angeles area, these statistics are highly suspect. It is possible that some teachers may have counted experience with Title I preschool classes, or perhaps experience in Childrens Centers, as Head Start. With this type of error revealed in verifiable data, it can be safely assumed that the non-verifiable responses are equally unreliable.

For purposes of the present analysis, these discrepancies, which can be presumed to hold for control and experimental teachers alike, are actually irrelevant. There is certainly far more variability among teachers within treatments than between treatments, and even where there seem to be meaningful differences, the large variance makes it highly unlikely that any of these are significantly greater than chance.

However, even if there had been significant between-group differences, numerous attempts to relate teacher characteristics to student performance have been repeatedly unsuccessful in establishing any strong correlations. Hopefully, the national evaluation, with its more extensive resources, will be able to apply the complex and sophisticated statistical procedures to extract from the tremendous pool of data some new insights into relevant teacher characteristics.

Post Observation of Teachers (P.O.T.)

The P.O.T. was developed after several conferences of the group of Center Directors concerned with the assessment of classroom and teacher variables, meeting with the Evaluation Coordinator. This instrument was intended to get at some of the more intangible ways teachers affect how and what children learn. It is in more than one sense a potpourri of items, with several different rating systems. In its attempt to reflect behaviors related to specific interventions of a number of Evaluation and Research Centers, it covers the range from coping styles and reinforcement behaviors to specific structured learning experiences. Assessing the data obtained with this instrument becomes an understandably difficult enterprise.

Certain items seemed to this Center to be too difficult to interpret, especially because so many of the data collectors reported "no basis for judgment, occasion did not occur" even when a response could have been expected. Thus, only 14 of the 33 items on the P.O.T. are represented in Tables 18 and 19. The numbers in front of the items are the original numbers from the P.O.T. and the brief statements are worded so that the highest value on the rating scale reflects the greatest amount of the indicated behavior. Since the instrument was administered five times over the year, there are two types of statistics possible: first, the average behavior over the five observations is given on the tables as the mean; and second, the change from the first to the last observation is given a code value representing the direction of change. Both the rating scale and the interpretation of the coding are provided at the bottom of Table 19.

Inspection of the means for any one item across classes shows wide variation, yet there is no perceptible difference in the pattern of the first ten classes (the experimental group) compared to the last eight (the control). Although there are two experimental classes with mean scores somewhat below the others (low score indicating more of the desired behavior), the differences are not significant when scores over items and classes are averaged together by treatments. This is also true when comparing treatments across classes for each item. Most consistent was the finding that very few teachers were apt to "put on a show" for the observers, that is, there was little evidence of any attempt to act differently so as to make a good impression. One might have expected that there would be more "company behavior" on the part of the control teachers, who did not have a chance to become as familiar with the observers as did the experimental teachers, but evidently the numerous contacts necessitated by the full battery of tests and observations dispelled any sense of strangeness that might have existed. On the other hand, there was very little difference between the first and last observations regardless of the treatment group.

There were on the whole very few extreme ratings of either 1 or 5; most frequently the various categories of behaviors were considered to be in the middle range. In terms of direction of change, the same tendency was found, that is, few changes from extremely low to extremely high or extremely high to extremely low. For the experimental group, six items showed a slight tendency to move from less frequent to more frequent demonstrations of the desired behavior, four items showed no change, and four items showed slight decrement. Similarly for the control group, five items showed improvement, six items showed no change, and three items showed barely perceptible negative change.

Table 18

Post Observation of Teacher with D Indicating Direction of Pre-Post Change

Classes	1. Responds to Individuals		2. Relies on Individual Activities		3. Attends to Individuals		4. Child Chooses Activity		5. Teacher Directs Play		7. Uses Approval		11. Stresses Variety in Verbs	
	M	D	M	D	M	D	M	D	M	D	M	D	M	D
1	1.4	3	2.4	3	3.0	3	1.4	4	2.2	3	1.4	3	3.0	1
2	3.0	5	2.2	1	3.2	4	3.2	4	2.8	5	3.8	4	4.2	3
3	1.6	3	2.0	2	2.8	3	2.0	3	2.2	3	1.8	3	5.0	3
4	1.8	3	3.0	3	3.2	3	2.6	5	1.6	4	1.8	3	4.8	3
5	1.8	3	2.0	3	2.6	2	1.6	1	2.2	3	2.2	3	4.4	4
6	1.6	2	1.2	3	3.0	3	1.4	3	2.6	3	1.2	3	2.0	3
7	1.6	4	2.2	3	2.8	3	2.4	3	2.2	2	1.8	3	4.4	3
8	3.8	3	2.4	3	4.2	3	4.8	3	3.0	3	4.4	3	4.8	3
9	4.2	5	2.4	3	4.0	3	2.6	2	1.8	3	3.2	4	4.4	3
10	2.4	3	2.4	2	1.4	3	2.8	2	2.6	2	2.0	3	4.2	3
11	2.8	3	2.2	2	2.6	3	2.6	3	3.0	2	3.4	2	4.0	3
12	1.4	3	3.0	3	2.6	4	2.8	4	2.0	2	1.8	3	3.4	3
13	1.6	3	2.8	3	2.6	3	3.2	3	2.8	2	2.6	3	3.4	3
14	1.8	3	2.8	3	3.0	3	3.0	3	3.0	3	1.8	3	4.0	3
15	2.0	2	3.0	3	3.2	3	1.2	3	2.0	1	2.6	3	4.0	3
16	2.0	5	3.8	3	2.8	3	2.8	3	2.2	4	1.4	3	3.8	3
17	2.6	3	3.8	2	2.8	2	2.8	3	2.6	3	2.4	3	4.4	3
18	2.2	3	1.8	3	1.4	3	2.0	3	3.2	1	1.8	3	4.2	3

Table 18 (cont'd.)

Classes	12. Stresses Variety in Adjectives		18. Allows Children to Complete Activities		24. Uses Varied Modes of Communication		26. Varies the Environment		27. Tries to Impress Observers		28. Provides Reasons		29. Encourages Varied Use of Materials	
	M	D	M	D	M	D	M	D	M	D	M	D	M	D
1	2.2	1	1.8	3	1.2	3	2.4	1	4.6	4	2.4	3	3.4	4
2	3.8	3	3.4	4	3.6	2	2.6	2	3.6	2	4.4	4	3.6	2
3	4.0	3	4.0	3	2.4	3	2.0	4	4.8	3	3.2	3	3.8	3
4	3.8	3	3.0	3	3.0	2	2.0	5	3.6	1	2.6	4	3.6	4
5	4.4	4	3.0	2	2.8	3	2.6	3	5.0	3	3.4	3	4.0	2
6	1.2	3	2.6	3	2.0	4	1.8	3	5.0	3	1.0	3	2.0	1
7	3.6	3	2.8	3	2.4	3	1.8	3	4.8	3	1.6	3	2.2	3
8	4.2	4	4.4	3	4.8	3	3.4	3	5.0	3	4.4	3	4.8	3
9	4.4	3	3.4	4	4.4	3	3.4	3	5.0	3	4.8	3	4.8	3
10	3.8	3	4.4	3	4.0	3	1.8	3	5.0	3	4.2	3	4.2	3
11	4.4	3	3.4	3	3.8	3	2.8	3	5.0	3	3.0	3	4.0	2
12	2.0	2	2.6	3	2.2	3	3.4	2	5.0	3	1.6	3	3.2	3
13	2.2	2	3.4	2	2.2	4	3.8	4	4.6	3	3.2	3	3.0	2
14	3.8	3	3.2	3	2.6	2	2.8	2	3.8	5	2.6	3	3.8	1
15	4.0	3	2.4	3	3.0	2	3.2	3	4.6	3	2.6	3	3.0	3
16	2.8	3	2.8	3	2.4	3	1.8	3	4.0	3	3.0	3	2.4	4
17	4.6	3	2.6	4	4.0	3	2.8	3	3.6	3	4.6	3	4.0	3
18	4.2	3	2.6	3	3.4	3	3.0	3	3.6	3	2.8	3	3.2	3

Table 19

Means and Standard Deviations for Experimental and Control Groups with Description of P.O.T. Items Used in Analysis

Item	Experimental				Control			
	M		D		M		D	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1. Responds to Individuals	2.3	1.0	3.4	1.0	2.1	0.5	3.1	0.7
2. Relies on Individual Activities	2.2	0.5	2.7	0.6	2.8	0.7	2.7	0.5
3. Attends to Individuals	3.0	0.7	2.9	0.4	2.6	0.5	3.0	0.5
4. Child Chooses Activity	2.3	0.9	2.8	1.1	2.6	0.6	3.1	0.4
5. Teacher Directs Play	2.2	0.4	3.0	0.7	2.6	0.4	2.0	0.9
7. Uses Approval	2.3	0.9	3.2	0.4	2.3	0.6	2.8	0.4
11. Stresses Variety in Verbs	4.1	0.9	2.9	0.7	3.9	0.4	3.0	0.0
12. Stresses Variety in Adjectives	3.5	1.1	3.0	0.7	3.5	1.0	2.7	0.5
18. Children Complete Activities	3.2	0.7	3.1	0.6	2.9	0.4	3.0	0.5
24. Varies Modes of Communication	3.0	1.1	3.0	0.5	3.0	0.7	3.0	0.6
26. Varies the Environment	2.4	0.6	3.0	1.0	3.0	0.5	3.0	0.6
27. Tries to Impress Observers	4.7	0.5	2.8	0.7	4.3	0.6	3.1	0.5
28. Provides Reasons	3.1	1.3	3.2	0.4	3.0	0.8	3.0	0.0
29. Encourages Variety of Materials	3.2	0.8	2.8	0.9	3.2	0.6	2.7	0.8

Rating Scale:

M = the mean rating per teacher for the five observations over the academic year.

Items 2, 3, and 4 are rated from 1 to 5 in terms of the greatest amount of the particular behavior described. Thus, for example, a rating of 1 for item 3 would mean that the teacher spends most of her time working with individual children, and 5 would mean she spends most of her time with large groups and little time with individual children.

For all other items, a simple frequency of occurrence scale is used such that:

1 = Constantly
2 = Frequently
3 = Occasionally
4 = Infrequently
5 = Opportunity present; behavior did not occur

Direction:

D = the slope of the line between the rating of the first and last observation for each teacher, using the following scale:

1 = Strong change from low to high
2 = Slight change from low to high
3 = No change
4 = Slight change from high to low
5 = Strong change from high to low

Observation of Substantive Curricular Interaction (OSCI)

It is expected that this instrument will eventually provide a great deal of information about classroom interactions and processes which will make it possible to relate the changes in children to the experiences to which they have been exposed. Unfortunately, the complete analysis of this data was not available in time for this report; furthermore, the type of analytic statistics required were beyond the resources of this center, both in terms of funds and expertise. As a compromise, a technique for the hand analysis of the data into meaningful albeit crude categories was devised. The method used is described in detail in the Appendix on pages 120 to 122.

Table 20a presents the gross descriptions of the classrooms in terms of how frequently activities were located indoors or outdoors, how frequently children initiated activities compared to teacher initiations, and the frequency of groups of various sizes. It also includes the frequency for the major categories of context and content codes.

Again it can be seen that there are more differences among agencies than there are between treatments within the same agency. Only in Agency III is there a notable between-treatment difference, with a larger percentage of children engaged in indoor activities under adult direction and relatively few children engaged in individual pursuits, compared to the control group which spent an almost equal amount of time indoors and outdoors but was very seldom under direct teacher supervision. All groups showed a preponderance of individual or small group activity with children and adults sharing locus of control on an almost 50-50 basis, except for Agency IV, where the teacher was in control almost 2/3 of the time. This agency had a considerable number of children in larger groups, which is consistent with the fact that IV B was the only group where programmed instruction was observed. Also consistent with this is the unusually high emphasis on cognitive content, 22.9% compared to the next highest cognitive input observed with Group I B of 13.5%.

As could be expected, routines took up a good portion of the day for each class, ranging from 13.6% to 40.1%. Comparatively little time was spent in passive watching, a category which also included listening to stories and music. In almost every class, the art activities took up more of the class time (approximately 20-25%) than any other activity except routines. The emphasis on structured cognitive input at Site IV B seems to have been at the expense of the large muscle activity, which showed 7.2% frequency as compared to that of the next frequency of 14.7% in I B, the second highest cognitive input class. The greatest inter-class variability was in the category of small muscle-perceptual discrimination activities. Here the range was from 5.7% to 23.5%. Both III A and III B were particularly low in this type of activity, as were II B and IV A.

Table 20b reports the data from the teacher segments of the OSCI. The same categories of inputs are examined for the Head Teacher, the Assistant Teacher, and then both teachers combined for each site. It

Table 20a

Hand Analysis of OSCI Data for Classroom Observations by Agency and Treatment

Description of Variables	% Frequency of Occurrence for each Treatment, by Agency											
	I A	I B	Mean	II A	II B	Mean	III A	III B	Mean	IV A	IV B	Mean
Location												
Indoor	79.6	76.3	78.0	53.7	61.3	57.5	71.2	42.5	56.9	76.5	82.4	79.5
Outdoors	20.4	23.7	22.1	47.3	38.7	42.5	28.8	57.5	43.2	23.5	17.6	20.6
Locus of Control												
Child	46.9	41.7	44.3	46.7	43.6	45.2	37.0	72.1	54.6	40.3	36.2	38.3
Adult	53.1	58.4	55.8	53.4	56.4	54.9	62.0	27.9	45.0	59.7	63.8	61.8
Group Size												
1 (Individual)	32.8	38.9	35.9	40.9	29.9	35.4	18.0	48.0	33.0	25.6	41.5	33.6
2-5	49.8	47.7	48.8	48.0	48.2	48.1	45.3	41.8	43.6	37.0	39.7	38.4
6-10	14.3	9.8	12.1	6.6	16.5	11.6	28.4	7.2	17.8	28.3	14.2	21.3
11 +	3.2	3.8	3.5	4.8	6.4	5.6	8.4	1.9	5.2	10.3	4.7	7.5
% Frequency for Context and Content Codes												
Routines	24.2	26.6	25.4	20.7	31.2	26.0	33.3	13.6	23.5	40.1	16.8	28.5
Watching	5.6	5.8	5.7	7.3	3.2	5.3	3.6	7.0	5.3	4.5	6.5	5.5
Cognitive	12.4	13.5	13.0	11.5	4.5	8.0	9.7	5.2	7.5	8.0	22.9	15.5
Arts	22.5	23.3	22.9	17.1	27.5	22.3	26.6	25.6	26.1	21.4	24.0	22.7
Small Muscle-Perceptual	19.0	19.4	19.2	23.5	8.4	16.0	5.7	9.0	7.4	8.1	17.1	12.6
Large Muscle	15.1	14.7	14.9	18.7	16.8	17.8	16.7	24.0	20.4	14.6	7.2	10.9

Table 20b

Hand Analysis of OSCI Data for Teacher Input by Agency and Treatment

Description of Variables	% Frequency of Occurrence for each Treatment, by Agency											
	I A	I B	Mean	II A	II B	Mean	III A	III B	Mean	IV A	IV B	Mean
Head Teacher												
Socialization	42.3	49.8	46.1	55.9	44.7	50.3	43.3	35.4	29.4	27.2	50.3	38.8
Cognitive	12.8	12.9	12.9	22.9	12.0	17.5	10.9	6.3	8.6	12.7	17.5	15.1
Sensory	5.5	9.2	7.4	3.3	4.6	4.0	2.8	4.2	3.5	1.4	6.8	4.1
Skill	1.8	0.6	1.2	3.7	2.3	3.0	4.2	0.7	2.5	1.7	3.5	2.6
Rules	3.1	2.9	3.0	2.3	3.6	3.0	4.4	9.0	6.7	2.7	4.2	3.5
No Input	39.0	33.1	36.1	27.5	40.6	34.1	38.6	57.6	48.1	52.8	32.4	42.6
Assistant Teacher												
Socialization	42.5	47.5	45.0	48.3	35.6	42.0	37.0	32.1	34.6	37.1	37.8	37.5
Cognitive	8.2	6.2	7.2	16.3	6.0	11.2	3.9	2.3	3.1	9.1	7.0	8.1
Sensory	5.5	4.5	5.0	3.2	2.5	2.9	0.2	1.8	1.0	3.7	10.2	7.0
Skill	2.0	0.7	1.4	3.6	3.7	3.7	4.0	1.1	2.6	15.1	4.8	10.0
Rules	3.1	5.3	4.2	2.8	2.6	2.7	1.9	2.8	2.4	1.1	1.3	1.2
No Input	40.2	34.1	37.2	37.5	55.2	46.4	49.1	61.0	55.1	44.5	50.5	47.5
Both Teachers Combined												
Socialization	42.4	48.7	45.6	52.1	40.2	46.2	40.2	33.8	37.0	32.1	44.0	38.1
Cognitive	10.5	9.6	10.1	19.6	9.0	14.3	7.4	4.3	5.9	10.9	12.3	11.6
Sensory	5.5	6.8	6.2	3.3	3.6	3.5	1.5	3.0	2.3	2.6	8.5	5.6
Skill	1.9	0.7	1.3	3.7	3.0	3.4	4.2	0.9	2.6	8.4	4.2	6.3
Rules	3.1	4.1	3.6	2.6	3.1	2.9	3.2	5.9	4.6	1.9	2.8	2.4
No Input	39.6	33.6	36.6	27.5	47.9	37.7	43.9	59.3	51.6	48.7	41.5	45.1

should be noted that the OSCI coding form permits more than one input notation for each teacher record, thus totals over 100% frequency are possible. Consistency across treatments compared to differences within agencies is again evident here; e.g., the Head Teacher was observed in the category described as "socialization" with a mean frequency of 29.4% for Agency III and 50.3% for Agency II. This category, which includes casual conversation and physical contact, occupies teachers more frequently than any other activity for all sites except III B and IV A, where "no input" has a frequency of 57.6% and 52.8% respectively. Cognitive input occupies approximately 12% of teacher time at all but three sites; there are two comparatively high cognitive input sites, II A and IV B (22.9% and 17.5%, respectively) and one comparatively low, III B (6.3%). "No input" has the second highest frequency for all classes except the two where it has the highest frequency.

The pattern for the assistant teachers is approximately the same as that of the Head Teachers, with the assistant teacher in III B showing the highest percentage of "no input" (61.0%). The only major departure from this pattern is in terms of the cognitive input category. Here the assistant teachers show generally low frequency, except for II A (16.3%). This is even more surprising inasmuch as this site had the highest frequency of cognitive input by the Head Teacher as well (22.9%).

Averaging over the two teachers at each site tends to flatten some of the differences, but the pattern remains quite distinctively the same: high frequency of socialization, lots of caretaking, with no input, a scattering of sensory, skill, and rules, and the relatively greatest variation among sites in terms of cognitive input.

Looking again at the teacher segments of the OSCI, Table 20c shows that at four sites almost one-third of the time the teachers are not in contact with any children. At sites II B and III A the frequency of non-involvement with children rises to 36%, and at sites III B and IV A the frequency of teachers observed without children is over 50% (51.4% and 52.1%, respectively). Where teachers are observed with children, the group size is most likely to be between two and ten children, although there is an almost equal possibility that the group will be one to eleven children.

The type of teacher involvement, shown on Table 20c, is consistent with the teacher input described in Table 20b. Thus, where there was high frequency of "no input," as in III B and IV A, there was also low frequency of active teacher participation and high frequency of teacher absence. It is reasonable to assume that where the Head Teacher is absent the Assistant Teacher will step in and take over. Surprisingly enough, the most involved Assistant Teacher was in IV B, a class which also had the most involved Head Teacher. On the whole, the assistants were almost as frequently involved with children as the Head Teacher, but less likely to be socializing with other adults or absent from the classroom.

The materials being used during each of the OSCI notations were grouped under eight categories and the combined data are presented in

Table 20c

Hand Analysis of OSCI Data for Involvement by Agency and Treatment

Description of Variables		% Frequency of Occurrence for each Treatment, by Agency											
		I A	I B	Mean	II A	II B	Mean	III A	III B	Mean	IV A	IV B	Mean
Size of Group of Teacher Involved													
1		16.1	19.4	17.8	16.1	17.0	16.5	9.7	17.7	13.7	11.3	23.3	17.3
2-5		21.2	13.7	17.5	29.6	19.6	24.6	18.7	16.0	17.4	8.8	17.9	13.4
6-10		23.7	20.8	22.3	10.9	14.5	12.7	20.3	10.8	15.6	18.4	17.4	17.9
11 +		9.5	18.0	13.8	15.3	13.9	14.6	15.9	5.4	10.7	9.7	11.6	10.7
None		28.9	28.3	28.6	29.5	36.6	33.1	35.7	51.4	43.6	52.1	29.7	40.9
Type of Teacher Involvement													
Head Teacher[a]	1	49.0	52.5	50.8	59.4	55.6	57.5	51.9	32.3	42.1	30.2	62.0	46.1
	2	19.0	16.4	17.7	11.3	9.9	10.6	16.7	19.0	17.9	15.8	9.6	12.7
	3	24.3	23.6	24.0	17.0	17.2	17.1	25.7	19.1	22.4	32.4	21.3	26.9
	4+5	8.0	10.1	9.1	12.8	19.5	16.2	6.1	28.1	17.1	22.6	7.1	14.9
Assistant Teacher[a]	1	47.6	44.2	45.9	46.3	40.2	43.3	38.4	28.5	33.5	44.5	48.1	46.3
	2	25.8	23.7	24.8	20.7	18.3	19.5	19.0	14.2	16.6	9.9	17.3	13.6
	3	21.3	22.8	22.1	20.8	36.0	28.4	37.0	35.1	36.1	25.8	30.2	28.0
	4+5	5.9	8.9	7.4	13.3	6.2	9.8	6.2	21.9	14.1	20.1	4.3	12.2
Both Teachers Combined[a]	1	48.3	48.4	48.4	53.1	47.9	50.5	45.1	30.4	37.8	37.4	55.0	46.2
	2	22.4	20.1	21.3	16.0	14.1	15.1	17.9	16.6	17.3	12.9	13.4	13.2
	3	22.8	23.2	23.0	18.9	26.6	22.8	31.4	27.2	29.3	29.1	25.8	27.5
	4+5	7.0	9.5	8.3	13.1	12.9	13.0	6.2	25.0	15.6	21.4	5.7	13.6

[a] 1 = Directly involved with children 3 = Involved in program activities, not children
2 = Passively supervising children 4 = Present visiting with other teachers; 5 = Not present

Table 21. Materials for art, dramatic play, large muscle, and small muscle activities are found most frequently in all of the classes. In contrast, music and dance, science, and language lessons have relatively meager material support, and programmed instruction was found in only two of the 18 classes. Both the highest (40) and lowest (22) number of available materials were found in the experimental classes, with no appreciable mean difference between the two treatments.

Observer's Report on the Class

At the end of the year, the observers, examiners, data aides, and all other evaluation personnel who had had any appreciable contact with a particular class separately filled out the Observer's Report on the Class. They then met together and compared ratings for the various items. Where major discrepancies were found, the reasons for the ratings were discussed and a consensus reached to represent the overall judgment for the particular class on seven important variables. These are listed in Table 22, together with an explanation of the scoring system. A quick scanning of the means for the experimental and control group reveal little likelihood that any reliable differences based on treatment will be found.

In addition to the scaled ratings for each of the seven categories, the ORC has several open-ended questions with room for the observer to make additional comments. A brief resume of these comments for each class is presented in Table 23. It is of considerable interest to compare the subjective statements describing the teacher in this instrument with the data obtained from the records of frequency of specific behaviors in the objective, structured context of the OSCI reported in Table 20. For example, the teacher at site III A was given probably the most glowing evaluation by the observers, yet on the OSCI she received the lowest ratings for involvement and input. It would no doubt be possible to do more precise quantitative comparisons of the data from these two instruments, but this was not feasible with the limited resources available for writing this final evaluation report.

Table 21

Variety of Materials Used by Categories

Class	Art	Music & Dance	Dramatic Play	Large Muscle	Sensory Small Muscle	Science	Language	Programmed Material	Total
1	6	1	9	11	9	1	3	0	40
2	5	2	6	9	6	3	3	0	34
3	6	0	7	9	8	2	5	0	37
4	7	0	7	10	7	1	3	0	35
5	4	2	7	9	7	0	1	0	30
6	6	1	9	9	9	2	3	0	39
7	4	1	3	7	7	2	2	0	26
8	4	1	2	7	6	1	1	0	22
9	6	1	4	8	6	2	2	0	29
10	4	1	4	7	5	2	4	0	27
11	5	1	6	10	9	0	3	0	34
12	3	2	7	9	6	2	4	0	33
13	5	0	6	7	8	1	1	0	28
14	4	0	5	5	5	1	3	0	23
15	4	1	7	10	8	2	3	0	35
16	4	2	7	8	7	0	2	1	31
17	5	2	6	5	7	0	5	0	30
18	7	1	3	3	8	2	2	1	27

Experimental: Mean - 31.9
S.D. - 5.7

Control: Mean - 30.1
S.D. - 4.0

Table 22

Observer's Report on the Class

Agency	Treatment	Classes	Sensory	Language	Group & Social	Concept Development	Academic Skill	Self-Esteem	Motivational
I	A	1	2	2	1	3	3	1	3
I	A	2	2	3	1	3	3	3	3
I	A	3	2	3	2	2	2	1	2
I	A	4	4	3	3	3	3	2	4
II	A	5	3	4	2	3	4	3	3
II	A	6	1	3	1	1	4	1	1
III	A	7	2	1	2	3	3	2	2
III	A	8	4	5	4	3	4	4	5
IV	A	9	5	5	5	5	5	5	5
IV	A	10	4	4	5	4	3	3	4
I	B	11	2	1	2	1	3	3	3
I	B	12	3	3	2	2	2	1	2
II	B	13	2	1	1	1	2	4	3
II	B	14	2	2	1	2	1	3	3
III	B	15	3	2	3	3	3	2	3
IV	B	16	3	1	2	2	2	2	1
IV	B	17	4	5	4	4	4	5	5
IV	B	18	4	4	4	4	4	4	5
	A	Mean	3.0	3.3	2.7	3.1	3.5	2.6	3.2
		S.D.	1.3	1.2	1.5	1.1	.9	1.4	1.3
	B	Mean	2.9	2.4	2.5	2.4	2.8	3.1	3.3
		S.D.	.8	1.5	1.1	1.2	1.0	1.2	1.3

Rating Scale:

1 = There was a very strong emphasis on this factor (i.e., observed some structured activity relevant to this goal almost every visit to the classroom)

2 = There was moderately strong emphasis on this factor (i.e., observed some structured activity relevant to this goal on over 60% of the visits to the classroom)

3 = There was some emphasis on this factor (i.e., observed some structured activity relevant to this goal about 30-50% of the visits)

4 = There was very little emphasis on this factor (i.e., observed structured activity relevant to this goal on less than 20% of the visits to this class)

5 = There was no emphasis on this factor (i.e., cannot recall any structured activity directed to this goal during class visits)

Table 23

Observer's Report on the Class

Summary of Open-Ended Items

Class No.	(B) Goals other than those listed (C) Extent of stability or disruption of program (D) Informal description
1	(B) Getting ready for regular school. (C) None. (D) Well-organized. Some child control. Not too much concern with behavior and quietness. Teacher had very strong voice for control, but assertive rather than mean or yelling. Emphasis on individuality--self as person--equality.
2	(B) None. (C) Teacher change at Christmas. (D) Teacher used loud voice in unsuccessful attempt to maintain discipline. A lot of music and game learning.
3	(B) None. (C) Change of teachers. (D) Outside a lot. Indoors, teachers were concerned with quietness and good behavior.
4	(B) Acquire knowledge of outside world; cultural activities and trips. (C) Change of teacher. (D) First teacher had an academic approach and children seemed to be more involved. This teacher was also interested in the families of the children. Everything was done with a goal in mind although children were not punished for not achieving these goals. Second teacher did not seem to be concerned with children as individuals, but rather with the class as a whole. However, she seemed to have little control.
5	(B) Table manners, taking turns, art, self-discipline, attention to parents. (C) None. (D) Children had many opportunities to make independent decisions. Teacher allowed noisy activities. Assistant teacher very actively engaged with children. She was always there on time. Head teacher was very good when she applied herself, but often seemed to be preoccupied with other concerns.
6	(B) Emphasis on socialization; art activities, etc. (C) None. This was a particularly stable class. (D) Fine class; much meaningful stimulation. Teacher seems aware of the needs of children and always has interesting things for them to do.
7	(B) Prepare for discipline of regular school, lining up, etc. Accept responsibility. (C) None.

Table 23 (cont'd.)

7	(D)	This class was one of five occupying one large open area. Everything was shared and there was much noise in general but particularly for this class which was near the phone where the teacher was constantly interrupted. They had many activities (flower making, painting) and the children were kept busy, but the teacher said this was the most difficult class she had ever had. Much individual attention and positive feedback given to children. Children encouraged to take initiative in activities and conversations. Disruptive children were calmed by embracing and talking to them. The worst thing about this class was lack of adequate indoor space and quiet. The best thing was the teacher's enthusiasm and willingness to let children experiment.
8	(B)	Learning to sit and listen.
	(C)	None.
	(D)	This was another of the five classes in the same location as the preceding class, with equally high noise level. The teacher often spoke derogatorily about the children in their presence. Class was heavily teacher-controlled. Teacher a "grandmother" type, very concerned with good behavior, lost her temper easily and reprimanded children severely for spilling food; children required to finish all their meal before being given dessert. Teacher controlled children by banishment from the group and negative comments to child concerned. Except for rare exceptions, activities were teacher-controlled, structured, and routine. Lack of enthusiasm apparent among the children as well as lack of initiative. Teacher was at times apathetic and hostile to observers. Fortunately there was an assistant teacher who was very responsive to the children and there was a large outdoor playground.
9	(B)	None.
	(C)	None.
	(D)	Children were left pretty much on their own. The main emphasis seemed to be on keeping children busy instead of teaching. Almost all the time was spent in free or art activities, with the teacher present but not participating. Teachers seemed to take turns working, one teacher taking care of two classes on alternate days. The teacher often stood with arms crossed, mainly engaged in discipline and control and was never seen outside with the children.
10	(B)	Emotional growth.
	(C)	Class in continuous state of disruption because three classes shared two rooms and there was a great deal of friction.
	(D)	Class space is very cramped. Site smelly, damp and dirty. Excellent teacher deserving of a better environment. Hampered by Head Teacher with little or no interest in the children, lack of space, inadequate materials, and transportation problems. Children all came by bus, which was usually very late and often broke down completely.

Table 23 (cont'd.)

11	(B) None. (C) None. (D) Children encouraged to make own decisions. The teacher was warm and made the children and parents feel important. However, she was cold and hostile to observers and other outsiders.
12	(B) None. (C) Teacher change. (D) Stability attained quickly after arrival of new permanent Head Teacher. Teacher very aware of children as individuals and dealt with personality problems by having individual "therapy" sessions with children who were upset, discussing their nightmares, family situations, etc. There was much freedom of personal expression in art work, where the teacher provided basic materials and encouraged the children to use their imaginations. Rather cramped quarters.
13	(B) Self-confidence. (C) None. (D) Only class with male teacher. Gave reassurance when children were uncertain and insecure.
14	(B) Learning own names. (C) None. (D) This class was mostly controlled by the teacher when they were inside, but when outside, the children were allowed to decide for themselves although the teacher was there when they needed her. Many toys and other equipment but not available without special permission.
15	(B) Self-confidence and social adjustment. (C) None. (D) This class had a good balance of child and teacher control. Almost always there was more than one on-going activity, and children were encouraged to make their own choices. Teachers were not overly concerned with "good" behavior and quietness and had high tolerance for noise and mess. Teacher spoke in a calm, soft voice, and was relaxed in her movements. Order was maintained by talking to and redirecting disruptive children.
16	(B) Awareness of multi-racial values; emotional adjustment, and language development (Teacher taught Spanish to black children and English to Mexican-American children.) (C) None. (D) The children were given some opportunity to make decisions, although the teacher kept careful control. Concerned with "good" behavior, but used logical reasons for imposing sanctions. Discipline was through re-direction of activity and reasoning with children. Much warm physical contact between teacher-child and child-child. Teacher gave a great deal of positive feedback, real listening to the child, respect. Perhaps over emphasis on academic part of program. Much work with cause and effect relationships. Asked more questions than

Table 23 (cont'd.)

16		gave answers. Exceedingly warm, free-flowing atmosphere. Classroom alive with color and sound. Activities were orderly and goal-directed. Teacher's warmth and concern for the children were the major strength of this class.
17	(B)	Cleanliness, appropriate classroom behavior.
	(C)	None.
	(D)	This class was primarily teacher controlled and the main concern was teaching children how to behave in school. Little noise was tolerated and free play or socialization discouraged. Classroom extremely quiet. Extremely teacher-controlled. Very negative atmosphere. Teacher overly concerned with quiet--to the extreme (no talking at lunch table!). Children inhibited and listless. Teacher rarely involved directly with children. Often spoke negatively with other adults about children in children's presence. Teacher gives orders, not suggestions or directions. Teacher seems bored.
18	(B)	Socialization.
	(C)	None.
	(D)	Afternoon class but teacher also very much involved in morning. Much socializing between teachers in both classes with resultant ignoring of children. Teacher was warm to children, but overly concerned with proper behavior. Children were left on their own much of the time both outdoors as well as inside. Much interference from teachers of morning class. Shared space was major problem.

Two instruments, Teachers Attitudes Toward Evaluation (TATE) and Teachers Expectations for Achievement of Children in Head Start (TEACH), were developed at UCLA specifically to measure the effect of the feedback intervention. These have been written up separately and are included in Part IV of this Final Report, pages 152-163 and 164-173. The findings on these two instruments are briefly summarized here.

Teachers Attitudes Toward Evaluation

The analysis of the responses on the TATE pretest indicated that there were distinct patterns of favorable and unfavorable attitudes toward various aspects of the evaluation concept. It was quite clear that teachers had the most positive attitudes toward the activities which provided them with data about a) their children, with special emphasis on the social-emotional area; b) the relationship between their class and the rest of the Head Start program; and c) the involvement of parents in the education of their children.

It is interesting to note that the lowest ratings were given when the testing was perceived as non-responsive to teacher needs or involved removing children from the classroom. The personal threat felt by teachers in an evaluation sample class is evident in the ratings of items having to do directly with the teacher or with classroom observation, which was seen as having a bad effect on children whereas in actuality the children for the most part were completely undisturbed by the observers.

At approximately the mid-point of the year, the group which was receiving the feedback intervention seemed to have reached a high level of rapport with the research staff. It was felt that this enthusiasm might be an ephemeral phenomenon, and that it might diminish considerably with habituation, so that by the time the posttest measure was taken the effectiveness of the intervention would not be measurable. For this reason, it was decided to administer the TATE to the experimental group. Since the agenda for the feedback meeting was already full, the TATE booklets were broken up and reassembled so as to contain half of the attitude and half of the belief scale items. These were then compiled as alternate forms and randomly presented to the teachers in the experimental group.

This mid-term evaluation demonstrated that the items rated most positively in the pretest were not affected by the attitude changes which occurred at this time, and those items most closely related to the site variable reflected the greatest amount of change. There was an increase in unfavorable attitudes toward observers in classrooms, rating teachers, and effect on child behavior of observers and testers. Positive attitudes toward evaluation were found for increased use of films and classification of classroom activities. With respect to the belief ratings reflecting knowledge of the substantive characteristics of an evaluation program, there was increased awareness that testing of samples of children provides information about all children, and that evaluation was concerned with measuring changes in children.

These mid-year results indicate the beginning of a trend showing that after teachers had been informed as to the rationale of the measures and invited to express their feelings and criticisms about them, they tended to become more tolerant of evaluation processes which involved interrupting the normal classroom activities. However, there was a tendency to become increasingly negative toward the items initially rated most negatively.

Inspection of the relative rank order positions of the items based on the pre and posttest scores revealed that the teachers continued to rate the items having to do with the overall Head Start programs more positively than items having to do with site activities. There was increased correspondence between attitudes about various activities and the belief that these activities are part of the evaluation program, even though many of the characteristics were specific to the activities of the UCLA Head Start Evaluation and Research Center alone.

In general, the attitude changes for the teachers receiving feedback paralleled those for the total group. However, they became significantly more positive about the time spent by observers in the classroom, and did not become more negative toward rating the teacher's performance and the observer's effect on child behavior. These particular differences support the hypothesis that teachers given feedback become less threatened and more positive toward the Evaluation personnel and activities.

On the belief scale, the experimental group significantly changed their beliefs about evaluation on all the items on which the entire group changed. In addition, only the teachers receiving feedback significantly decreased their belief that Evaluation includes spending class time participating in Head Start research, and that Evaluation includes removing children from the classroom for testing, knowing how their class compares with other Head Start classes, and testing children. Thus while the total group of teachers did show some changes in their beliefs about Evaluation which in fact are congruent with the program, the experimental group changed their beliefs the most and in a direction consistent with the activities of the Evaluation program. Thus these results definitely tend to support the hypothesis that providing feedback to teachers reduces threat, increases their reception to the program, and helps them to distinguish fact from fiction insofar as specific program activities are concerned.

Teachers Expectations for Achievement of Children in Head Start (TEACH)

It was hypothesized that the teachers in the feedback group, compared to the controls, would show a higher degree of correspondence at the end of the year between their ratings of the importance of certain behavioral goals for Head Start children and their estimates of achievement of these goals. Also, there would be a tendency to place greater value on those goals they felt they had achieved compared to goals with which they had had little success.

According to the pretest data, the means on the Importance Scale were highest for Independence, Curiosity, Social Awareness, and Self-Awareness, in that order. They were decreasingly lower for Motor Skills, Sensory and Cognitive Skills, and School Appropriate Behavior. Developmental or readiness factors had higher rank orders than educational factors; the rank ordering on the factors for the three treatment groups showed the same trend as for the total group. These results indicate considerable face validity in terms of the fact that Head Start teachers apparently view the development of emotional, social, and situational maturity as more important than academic skills *per se*.

After the posttest data were obtained, a series of t-tests was run between the pre- and posttest scores on each item. There were no statistically-significant changes between the pretest-posttest Importance ratings. In other words, the teachers included in this sample had very stable feelings about their objectives. The speculation that there might be a decrement in the value placed upon an objective when it was not adequately achieved was not supported. The differences that did obtain were between the expectations for achievement and the ratings of level of achievement obtained. Without exception, these differences were in terms of a decrease in mean scores. Thus it appears that for the total group of teachers, the degree of correspondence between Importance and Expectation increased during the year.

Interestingly enough, the Expectation scores were generally higher than the Importance ratings. Several speculations may be advanced for these findings. The teachers may stress the behaviors for which they hold the highest importance, thus a type of self-fulfilling prophecy could be operative. If this is true, there should be lower Expectation ratings for those factors which are rated low on the Importance Scale. While the rank order of the Expectation ratings follows the same direction as the Importance ratings, the numerically higher Expectation ratings seem to contradict this assumption. Another possibility is that while the teachers do in fact view readiness for learning as a prime function of Head Start, they feel that this goal can best be achieved by first establishing strengths in such non-cognitive areas as basic trust and self-esteem.

While greater variance was found among the teachers in the experimental treatment, the number of items was so large and the sample size of teachers so small that it would have been meaningless to look for statistical differences between the pretest-posttest scores by treatment group.

Class Facilities and Resources Inventory

As indicated on page 33, this instrument is an exceedingly comprehensive assessment of the physical environment of the classroom. It was impossible to do justice to the wealth of data which it provided, and only a selected number of variables were combined and weighted for this analysis.

Although given as a pre-post instrument, only the post data are analyzed since a complete revision of the format was effected between the testing points and the items are thus not readily comparable. The changes were made because a great many objections had been raised at the Evaluation Coordinator's meeting. In many cases the items were found difficult because there were no anchor points. For example, how many is an "adequate" supply of books, or art materials, or whatever? Also, getting precise measurements of space either indoors or outdoors, proved to be quite unfeasible. Finally, the layout of the questions and the spaces for responding were confusing to the data collectors, many of whom were recruited from the local poverty communities. The revision was therefore directed toward making the items more objective and not dependent upon evaluative judgments nor comparisons. The entire emphasis was to report on the basis of what could be seen or counted. Even so, there were many problems in obtaining complete protocols from the various sites, and the data should be examined with some healthy scepticism. The items used in Tables 24 and 25 and the coding or rating system are presented on the next page.

According to Table 24, it can be seen that each of the control classes had a separate classroom to itself. In contrast, one experimental class had two rooms, two shared two rooms among three classes, and two shared one room with three classes. While it is impossible to determine whether or not this was a significant variable, it is clear that there was much more crowding in the experimental group.

The length of the school day also militated somewhat against the teachers in the feedback classes. All the control classes were three hours or longer, with two classes four and a half hours; two of the experimental classes were only two and a half hours long, four were three and a half, and four were four hours. In general, there were few distinguishing features between the experimental and control groups on any of the other items, with one important exception: the amount of available play equipment. In the comparison between experimental and control classes presented in Table 25 it can be seen that there is considerable difference in every area except large muscle equipment. However, on the other items, even where there were only small differences they were in favor of the control group. These findings, where both experimental and control classes were under the sponsorship of the same Delegate Agencies, point up most dramatically the need for caution when discussing "Head Start" as if it were a consistent, replicated program.

Explanation of Description and Coding of Class Facilities Items Used in the UCLA Analysis

1. Ratio of classrooms per class
   1 = 2   3 = .66
   2 = 1   4 = .33

3. Adequacy of outdoor space
   1 = Excellent   3 = Adequate
   2 = Good   4 = Poor

4. Availability of books
   1 = Sufficient; available; permission to take home
   2 = Sufficient; available; no permission to take home
   3 = Sufficient; unavailable; no permission to take home
   4 = Limited; available; no permission to take home

5. Overall indoor appearance
   1 = Excellent   3 = Average
   2 = Good   4 = Poor

6. Amount of play equipment:
   Each class was considered to have an item if it was seen and available for use and in any condition except very poor. The coding indicates the actual number of items of each category.

7. Overall rating of site
   1 = Good
   2 = Adequate
   3 = Poor

8. Transportation (method of getting to class)
   1 = Walks with adult or alone, less than 1/2 mile
   2 = Walks with adult 1/2 to 1 mile
   3 = Head Start bus

9. Quality of lighting
   1 = Very bright   3 = Adequate
   2 = Bright   4 = Dim

10. Storage facilities
    1 = Good
    2 = Average
    3 = Poor

11. Adequacy of special areas (housekeeping, art, etc.)
    1 = Many
    2 = Several
    3 = Few

Table 24

Class Facilities

Category	Class																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Ratio of classrooms per class	1	2	2	2	2	2	4	4	3	3	2	2	2	2	2	2	2	2
2. Number of hours per session	3.5	3.5	3.5	3.5	4	4	4	4	2.5	2.5	4.5	4.5	4	4	3.5	3	3	3
3. Adequacy of outdoor space	2	2	2	2	3	3	2	2	2	2	4	4	1	1	4	2	2	2
4. Availability of books	1	1	1	2	2	2	3	3	3	3	3	3	2	2	1	3	2	2
5. Overall indoor appearance	2	2	2	2	1	1	2	3	4	4	3	3	2	2	3	1	2	2
6. Amount of play equipment:																		
a) Large muscle	5	5	5	5	4	4	3	3	2	2	4	4	4	4	5	4	4	4
b) Pretend	6	6	5	6	5	5	3	3	1	1	5	5	5	4	6	6	6	4
c) Creative	6	6	5	3	7	7	3	3	3	3	5	7	5	5	6	7	7	6
d) Learning	4	5	6	4	5	5	4	4	3	3	3	4	6	6	7	8	8	8
7. Overall rating of site	2	2	2	2	2	2	3	3	3	3	2	2	2	2	1	2	2	2
8. Transportation	1	1	1	1	1	1	2	2	3	3	2	2	2	2	1	2	2	2
9. Quality of lighting	2	2	2	2	2	2	3	3	3	3	4	4	2	2	4	2	2	2
10. Storage facilities	3	3	3	3	1	1	2	2	2	2	3	2	2	2	3	3	3	3
11. Adequacy of special areas	3	2	2	2	1	1	2	2	3	3	3	3	2	2	3	1	1	1

Table 25

Class Facilities (Means and Standard Deviations for Experimental and Control Groups)

Category	Experimental		Control	
	Mean	S.D.	Mean	S.D.
1. Ratio of classrooms per class	2.5	0.9	2.0	0.0
2. Number of hours per session	3.5	0.6	3.7	0.6
3. Adequacy of outdoor space	2.3	0.4	2.5	1.2
4. Availability of books	2.2	0.8	2.3	0.6
5. Overall indoor appearance	2.3	1.1	2.3	0.6
6. Amount of play equipment:				
a) Large muscle	3.7	1.2	4.1	0.3
b) Pretend	4.0	1.9	5.1	0.8
c) Creative	4.7	1.8	6.0	0.9
d) Learning	4.2	0.9	6.2	1.9
7. Overall rating of site	2.4	0.5	1.9	0.3
8. Transportation	1.6	0.8	1.9	0.3
9. Quality of lighting	2.4	0.5	2.8	1.0
10. Storage facilities	2.1	0.8	2.7	0.5
11. Adequacy of special areas	2.1	0.8	2.0	0.9

Correlational Analysis

The final section of this report will be devoted to an examination of some of the relationships among the various instruments used in the evaluation . In this correlational analysis, since there were so few large differences among treatments, the scores of the children in the experimental and control 1 classes were combined. The means and standard deviations, plus the number which identifies the variable in the correlation matrix, are presented in Table 26. It should be noted that the means for the class variables were computed by assigning the same value to all the children in the class. This produces a weighted mean which is not always identical to the mean for that variable when computed on a class basis. However, it would have been impossible to run a correlation using both individual child measures and classroom ratings without this procedure.

The correlation tables are presented in two parts. Part 1 is concerned with the relationships between individual child variables and all other measures. Only the correlations significant at either the .05 or .01 level are presented in Table 27. Because of the extremely restricted age range, it is not surprising that there are so few valid correlations between chronological age and the other variables. It is therefore somewhat surprising to find correlations between C.A. and posttest scores on the Caldwell and WPPSI. It is also interesting to find negative relationships between the age variable and the amount of teacher experience with children other than or older than Head Start.

One of the most unexpected findings from the correlation matrix is that there isn't a single reliable correlation between attendance and any of the posttest or gain scores. If the attendance variable had had a rather limited range, there might have been some reasonable explanation for this lack of measurable effect; however, there was a standard deviation of 21.6 from a mean of 143.3 days, with a population of 131 children, which would rule out this type of rationalization.

On the whole, the correlation between attendance and other variables tends to be positive, thus the negative correlations deserve special attention. Most of these make some sense. For instance, as the teacher increases her attention to individual children there is a drop in attendance. One interpretation might be that there is a focusing on specific individual children with a concomitant lessening of attention to all children, and those who are ignored may be dropping by the wayside. Another negative correlation is found with length of day. Evidently children who are enrolled in classes which have long daily sessions are not as apt to come as frequently as children in short daily sessions.

The posttest scores on the various individual child measures show strong intercorrelations, as do the posttest-gain correlations for the individual tests. However, there are few significant correlations between these scores and the large number of classroom and teacher variables.

Table 26

Means and Standard Deviations for All Variables for All Children

	N	Mean	SD
Chronological Age (1)[a]	128	52.2	3.6
Attendance in Days (6)	131	143.3	21.6

Individual Tests	Pretest			Posttest			Gain		
	N	M	SD	N	M	SD	N	M	SD
Caldwell (2, 7, 11)	128	33.1	10.7	125	42.2	9.4	122	9.4	8.1
Gumpgookies (2, 8, 12)	127	33.7	7.6	125	41.9	7.6	115	7.7	7.5
WPPSI (4, 9, 13)	130	10.1	3.4	125	11.6	3.3	123	1.6	3.4
Binet (5, 10, 14)	128	89.0	11.4	129	92.8	11.7	126	3.9	9.7

P.O.T. Items	Means		Direction	
	M	SD	M	SD
1. Responds to Individuals	(15) 2.2	0.8	(29) 3.3	0.9
2. Relies on Individual Activities	(16) 2.5	0.7	(30) 2.7	0.5
3. Attends to Individuals	(17) 2.8	0.7	(31) 2.9	0.5
4. Child Chooses Activity	(18) 2.4	0.8	(32) 2.9	0.9
5. Teacher Directs Play	(19) 2.4	0.5	(33) 2.6	1.0
7. Uses Approval	(20) 2.3	0.8	(34) 3.0	0.4
11. Stresses Variety in Verbs	(21) 4.0	0.7	(35) 3.0	0.5
12. Stresses Variety in Adjectives	(22) 3.5	1.0	(36) 2.9	0.7
18. Children Complete Activities	(23) 3.1	0.6	(37) 3.0	0.6
24. Varies Modes of Communication	(24) 3.0	0.9	(38) 3.0	0.5
26. Varies the Environment	(25) 2.7	0.7	(39) 3.0	0.8
27. Tries to Impress Observers	(26) 4.5	0.6	(40) 3.0	0.7
28. Provides Reasons	(27) 3.1	1.1	(41) 3.1	0.3
29. Encourages Variety of Materials	(28) 3.2	0.7	(42) 2.7	0.8

[a] Number in parentheses refers to variable number on correlation matrix.

Table 26 (cont'd.)

	Mean	S.D.
ORC Variables		
Sensory (43)	2.9	1.1
Language (44)	2.9	1.4
Group and Social (45)	2.6	1.4
Concept Development (46)	2.8	1.2
Academic Skill (47)	3.2	1.0
Self-Esteem (48)	2.8	1.4
Motivational (49)	3.2	1.3
Teacher Characteristics		
Age (50)	5.1	1.5
General Education (51)	3.9	1.3
Early Childhood Education (52)	2.8	1.3
Head Start Training (53)	5.4	2.7
Special Training (54)	4.5	3.0
Months of Special Training (55)	3.7	2.3
Teaching Experience:		
a) Children other than Head Start (56)	50.5	61.2
b) Head Start only (57)	38.5	13.3
c) Children older than Head Start (58)	23.6	29.8
d) Preschool children not Head Start (59)	26.8	40.1
e) Disadvantaged preschool children (60)	53.3	26.6
Class Facilities		
Classrooms per class (61)	2.3	0.7
Hours per day (62)	3.6	0.6
Outdoor space (63)	2.4	0.9
Books (64)	2.2	0.7
Indoor appearance (65)	2.3	0.9
Play equipment for:		
Large muscle (66)	3.9	0.9
Pretend (67)	4.5	1.6
Creative (68)	5.2	1.6
Learning (69)	5.1	1.7
Overall rating (70)	2.2	0.5
Transportation to class (71)	1.8	0.7
Quality of lighting (72)	2.6	0.8
Storage facilities (73)	2.3	0.7
Special areas (74)	2.0	0.8

Table 27

Significant Correlations[a] Among All Variables Based on Total Population Scores for the Major Variables Used in the UCLA Analyses

Chronological Age (1)	Attendance (6)	Caldwell Post (7)	Caldwell Gain (11)	Binet Post (10)	Binet Gain (14)
	15) .43	8) .40	61) -.21	14) .45	57) .22
	20) .21	9) .43	70) -.21	31) .21	
7) .26	24) .34	10) .52		33) -.21	
9) .30	27) .37	11) .22		47) -.28	
16) .26	28) .27	31) .28		53) -.20	
21) .24	29) .21	43) -.22		72) .20	
22) .22	32) -.21	44) -.20		74) .20	
24) .21	37) .25	46) -.28			
26) -.20	40) .20	47) -.25			
27) .21	43) .36	49) -.24			
42) .22	44) .22				
43) .27	45) .36	Gumpgookies		WPPSI	
45) .23	46) .34	Post (8)	Gain (12)	Post (9)	Gain (13)
52) -.20	47) .31				
56) -.31	48) .44				
58) -.22	49) .39	9) .29	21) .21	10) .34	38) .20
59) -.32	55) .21	10) .23	22) .23	13) .52	
60) -.26	62) -.40	12) .68	24) .24	58) .20	
73) .28	64) .21	48) -.25	35) .31	73) .24	
	65) .27	63) .21	36) .35		
	66) -.37	71) -.24	44) .25		
	67) -.29		59) -.20		
	70) .22		60) -.22		
	71) .40		74) -.22		

[a] For all correlations in this table:
.195 to .227 $p < .05$
.228+ $p < .01$

The second part of the matrix, Table 28, looks at the relationships between the latter set of measures. This is an extensive matrix table and hard to interpret without multivariate analysis; however, it is clear that there are far more reliable correlations than could be expected in a table this size. (It should be noted that the correlation matrix is presented in a rather unconventional format, so as to avoid the wastefulness of a "mirror-image" type of presentation. While it would have been infinitely desirable to present the entire table on one page, the amount of reduction needed would have made the data well-nigh unreadable.)

SUMMARY

For the most part, the discussion of the results has accompanied the presentation in the preceding section. To avoid further redundancy in an already over-long presentation, only a few closing remarks seem to be appropriate. First of all, it is abundantly clear that it is well-nigh impossible to pinpoint any particular set of circumstances as definitive ingredients of a successful Head Start program, especially where most of the individual classes offer quite similar fare. Secondly, it certainly cannot be shown from the present study that teachers who are given feedback about the areas of deficiency of specific children proceed to do anything meaningful with this information. It may be that this inference is unduly harsh, and that without feedback the children in the experimental group might have shown decrements in comparison with the control, perhaps due to the slightly unfavorable balance in almost all of the comparison variables. Or perhaps there were differences which were not susceptible to measurement. Whatever the cause or combination of causes, the statistical data fail to confirm any reliable differences in terms of performance of children on the standard measures used in the evaluation.

With reference to attitudes of teachers, the intervention was far more effective. Although there is some evidence of this in the TEACH and the TATE, the most convincing and gratifying proof of the value of the feedback intervention is in the consistently good rapport which was established and maintained in the local community. Whereas in many parts of the country evaluation teams are having great difficulty obtaining entry into classes in this area those agencies who have been most frequently and intensively involved in evaluation studies are most receptive to new research endeavors. That maintaining open channels of communication with teachers while carrying on evaluation activities can produce increased understanding and rapport is indeed a worthwhile finding of this investigation.

Table 28

Correlation Matrix for All Class Variables
(Significant at .05 or .01 level)

	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
15	.69	.55	.62	.72	.66	.28	-	-	-	-	-	-.36	-	-.20	-.41
16	-	-	-.38	.22	-	-	-	-.31	-	-.23	-	-	.20	-.25	-
17	-	-	.25	-	-	.42	-	.20	-	-	-	-	-	.20	-
18	.23	-	-	.46	.33	.51	-	-	.20	-.34	-	-	-	-	-.25
19	-	-.30	-	.29	-	-	-	.27	-	-	.26	-	-.20	-	-
20	.33	-	.29	.61	.51	.32	.25	.39	-	-	-	-.30	-	-	-.42
21	.55	.53	-	.42	.55	-	-.29	-	-	-.56	-.24	-	.30	-	-.25
22	.62	.59	.36	.60	.67	-	-.27	-	-	-.28	-	-.20	.21	-	-.28
23	.33	-	-	.32	.23	.50	-	-	-	-.28	-	-	-	-	-.34
24	.81	.56	.54	.76	.75	.21	-	-	-	-	-	-.38	-	-	-.47
25	[illegible]	-	-	.57	.49	-	.31	.25	-	.28	.73	-.33	-.36	-	-.40
26	-	-.22	-	-.22	-.31	-	.36	.27	-	.42	-	-	-.39	-	-
27	.68	.60	.30	.79	.73	.30	-.31	-	-	-.20	-	-.53	-	-.35	-.55
28	.56	.43	.40	.63	.71	-	-	-	-	-	.20	-.45	-.24	-	-.58
29	.27	.43	-	.36	-	.35	-.58	-.32	-	-	-	-.21	-	-.26	-
30	-	-	-	-	-	-	-	-	.25	-	-	.33	-.21	.26	.31
31	-	-.22	-.40	-.38	-.27	-	.34	-	-.43	.29	-	-	-	-	-
32	-.26	-.23	-.34	-.41	-	-	.33	-	-.24	-	-	-	-	-.23	-
33	-	-	-	-	-	.45	-.28	-	-	-.30	-.48	-.27	.38	-	-.28
34	.34	.62	.30	.31	.29	.25	-.45	-.40	-	-	-	-	-	-.22	-
35	-	-	-	.29	-	-	-	-	.46	-.30	-	-	-	.27	-.31
36	.35	.20	.41	.30	-	.39	-	-	.48	-.36	-.44	-	.22	.36	-.20
37	.51	.55	.39	.30	.39	-	-.22	-.48	-	-	-	-	-	-.30	-
38	-.22	-.30	-	-	-.26	-	-	-	-	-	-	-	-.46	-	-
39	-	-	-	-	-	-	-	.24	.40	-.52	-	-.20	.40	-	-.27
40	-.20	-	-	-	-	-	-.23	-	-	.43	.46	.43	-.38	.34	.40
41	-	-	-	-	-	.21	-	-	-	-.28	-.25	-	.46	-	-
42	.40	.49	-	-	.27	-	-.32	-.40	-.37	-.33	-.31	-.31	.28	-.52	-
43	.91	.84	.51	.65	.81	-	-.23	-.50	-	-	-	-.47	-	-.37	-.44
44	.71	.72	.71	.49	.67	-	-	-.52	-	-	-	-.38	-.20	-	-.48

Table 28 (cont'd.)

60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	
-.41	.45	-.47	-	.41	.64	-.61	-.58	-.45	-.29	.52	.60	.28	-	.27	15
-	-	-.21	-	-	-	-	.28	-	.33	-	.22	-	.47	-	16
-	.26	-	-	-	-	-	-	-.29	-.33	.20	-	-	-	.25	17
-.34	.48	-	-.38	.50	.35	-.36	-.27	-.43	-	.47	.51	-	-	-	18
-	-	-	-	-	-	-	-	.23	.31	-	-	-	-	-.28	19
-.50	.30	-	-	.23	.54	-.29	-.28	-.38	-.30	.26	.33	.41	-	.35	20
-.39	.41	-.30	-.20	-	.30	-	-.25	-.49	-	.28	.22	-	.35	-	21
-.42	.30	-.35	-	-	.35	-.22	-.28	-.33	-	-	.26	-	.31	-	22
-.39	.52	-	-.23	.35	.44	-.46	-.53	-.57	-.35	.51	.42	-	-	-	23
-.60	.49	-.43	-	.43	.61	-.55	-.52	-.43	-	.43	.58	.30	-	-	24
-.58	-	-	-	-	.42	-	-	-	-	-	.33	.29	-	.29	25
-	.31	.32	.42	.34	.30	-.42	-.42	-	-.65	.33	-	.50	-.59	.44	26
-.69	.20	-.50	-.33	-	.46	-.40	-.36	-.30	-	.34	.48	-	.20	-	27
-.70	-	-.21	-	-	.53	-.32	-.31	-	-.28	.20	.36	.24	-	.29	28
-	.36	-.45	-.35	.39	.24	-.44	-.39	-.34	-	.53	.50	-	-	-	29
.20	-	-	-	-	-	-	-	-	-	-	-	-	-.39	-	30
-	-	-	-	-	.34	-	-	-	-.29	-	-	.38	-	.50	31
.21	-.26	-	-	-.22	-	.54	.49	-	-	-.26	-.31	-	.57	-	32
-	-	-	-	-	-.27	-	.22	-	-	-	-.27	-.40	-	-.20	33
-	.24	-.60	-.41	-	.22	-.34	-.41	-.22	-	.36	.23	-.24	-	-	34
-.60	.30	-	-	.29	-	-.22	-	-	-	-	-	-	-.45	-.41	35
-.27	.46	-	-	.24	-	-.27	-.23	-	-	-	-	-	-.31	-.43	36
-	-	-.56	-	-	.46	-.26	-.25	-.20	-	.28	.36	-	.39	.21	37
-	-	-	-	.21	-.22	-.27	-	.20	-	-	-	-	-.49	-.25	38
-.23	-	-	-.20	-	-	-	-	-.36	-	-	-	-	-	-.24	39
.30	-	-	-	-	-	-	-	.24	-	-	-	-	-	-	40
-	-	-	-	-.24	-	.38	.29	-.22	-	-	-.34	-.22	.28	-	41
-	-	-.47	-	-	.23	-	-	-.26	-	-	-	-	.60	.23	42
-.56	.30	-.69	-	.33	.54	-.48	-.48	-.37	-	.37	.56	-	-	-	43
-.58	-	-.53	-	-	.27	-.38	-.39	-	-	.32	.23	-	-	-	44

Table 28 (cont'd.)

	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	
		.82	.59	.62	.78	-	-	-.47	-	-	-	-.42	-	-.36	-.36	45
16	-		.62	.53	.74	-	-.50	-.56	-	-	-	-.26	-	-.34	-	46
17	.48	-		.49	.54	-	-	-	.26	-	-.20	-.32	-	-.24	-.32	47
18	.45	.39	.24		.81	-	-	-	.38	-	.37	-.38	-	-.22	-.42	48
19	-	-.24	-.45	.27		-	-	-	-	-	.21	-.40	-	-.28	-.40	49
20	.78	-	.45	.60	.22		-.27	-	-	-.32	-.23	.23	.47	.31	-	50
21	.40	.28	-	.36	-	.42		.50	-	.39	-	-	-.32	-	-	51
22	.57	-	-	-	-	.53	.87		.32	-	.33	.20	-	.27	-	52
23	.46	-	-	.62	.20	.54	.48	.35		-.24	-	.24	-	.42	-	53
24	.88	-	-	.50	-	.74	.59	.75	.56		.45	-	-.37	-	-	54
25	.38	.21	.21	.35	.20	.59	-	-	-	.31		-	-.25	-	-	55
26	-	-.43	.22	-	-.20	-	-.23	-.20	.29	-	-		-	.83	.91	56
27	.76	.32	.23	.47	-	.67	.60	.67	.53	.77	.38	-		-	-	57
28	.61	-	-	.27	-	.64	.49	.67	.43	.71	.46	-	.77		.53	58
29	.53	.25	.32	.35	-.29	.29	.36	.27	-	.31	-	-	.47	-		59
30	-.22	-	-	-.20	-.31	-.35	-.25	-.34	-.38	-.36	-	-	-.41	-.33	-	60
31	-	-	-	.23	-	-	-.24	-.43	-	-	-	-	-.31	-.38	-	61
32	-.28	.25	-	-	-	-	-	-.37	-.31	-.29	-	-.41	-.36	-.42	-	62
33	-	-	.47	-	-.37	-	-	-	-	-	-.35	-	.26	-	.43	63
34	.43	-	.44	-	-.40	-	-	-	-	.21	-	-	.41	-	.59	64
35	-	-	-	-	-	.20	.35	.38	.39	.36	-	-	-	.21	-	65
36	.37	-	-	-	-	.32	.50	.60	.44	.57	-	-	.30	.29	-	66
37	.62	.22	.37	-	-	.25	-	.30	-	.51	-	-.27	.40	-	.45	67
38	-	-.33	-	-	.28	-	-.57	-.54	-	-.26	-	.44	-	-	-	68
39	-	-	-	-	-	-	.38	.20	.41	.20	-	-	-	-	-	69
40	-	-	-	-	.33	-	-.31	-	-	-.23	-	.24	-	-	-	70
41	-	-	-	-	-.24	-	.25	-	-	-	-.21	-.50	-	-	-	71
42	-	.51	-	-	-.48	-	.42	-	-	-	-	-.28	-	-	.34	72
43	.64	.35	-	.29	-.24	.32	.59	.60	.22	.72	.33	-.22	.68	.57	.38	73
44	.55	-	-	-	-	.23	.32	.42	.21	.60	-	-	.57	.57	-	
	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	

Table 28 (cont'd.)

	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	
	-.54	.45	-.73	-	.38	.65	-.62	-.63	-.45	-	.47	.64	.22	-	-	45
16	-.26	.33	-.79	-.24	-	.44	-.46	-.52	-.34	-	.44	.42	-	-	-	46
17	-.29	.27	-.42	-	-	-	-.44	-.40	-	-	.35	-	-	-.25	-	47
18	-.53	.29	-.46	-.37	.28	.37	-.52	-.43	-.26	-	.34	.62	-	-	-	48
19	-.53	.20	-.54	-.29	-	.52	-.34	-.40	-.39	-	.30	.46	-	.26	-	49
20	.30	.37	-	-.34	-	-	-	-	-.25	-	.26	-	-	-	-	50
21	-.33	-	.60	.51	-	.29	-	-	-	-.40	-.20	-	.55	-	.41	51
22	-	-	.59	-	-.22	-	.29	.21	-	-.21	-.24	-.33	-	-	-	52
23	-	.26	.24	-	.38	-.34	-.20	-	-	-	.20	-	-.22	-.31	-.48	53
24	-	-.29	-	.49	-	.38	-	-	-	-.38	-	-	.48	-	.55	54
25	-	-.34	-	-	-.20	-	-	-	-	-	-.29	.20	-	-	.26	55
26	.67	.24	.40	-	-	-.25	-	-	-	-	-	-.20	-	-.30	-	56
27	.39	-	-.30	-	-.45	-.22	.43	.34	-	.33	-.34	-.39	-.26	.47	-	57
28	.31	.26	.55	-	-	-.27	-	-	-	-	-	-.20	-	-.48	-	58
29	.79	-	-	-	-	-	-	-	-	-	-	-	-	-	-	59
30		-	-	-.26	-.35	-.47	.31	.26	-	-	-	-.41	-.36	-	-	60
31	-		-	-	.62	.40	-.73	-.72	-.69	-.36	.80	.53	.32	-.28	-	61
32	-	.55		.44	-	-.34	.39	.43	.26	-	-.31	-.45	.26	-.27	-	62
33	-.26	-	.22		-	-	-	.23	.36	-.22	-.37	-.25	.68	-	.25	63
34	-	-	-	.37		.40	-.77	-.59	-.39	-.42	.66	.75	.47	-.26	-	64
35	-	-.31	-.54	-	-		-.57	-.66	-.62	-.52	.45	.70	.68	-	.81	65
36	-	-.44	-.54	-	-	.81		.92	.54	.42	-.85	-.86	-.28	.42	-.21	66
37	-.39	-	.20	.26	.47	-.25	-		.68	.49	-.84	-.78	-.27	.34	-.34	67
38	.20	-	-.22	-	-	-	-.20	-.32		.58	-.71	-.49	-.24	-	-.49	68
39	-	-.22	-	-	-	.49	.37	-.22	-		-.56	-.23	-.47	.31	-.67	69
40	-	-	-.43	-.48	-	-.31	-.26	-	.21	-.67		.67	-	-.30	-	70
41	-.26	.27	.57	-	.22	-	-	-	-.54	.34	-.75		.35	-	.34	71
42	-	-	.43	.22	-	-.42	-.27	.26	-.38	-	-.28	.20		-	.72	72
43	-	-	-	-	.44	-	.28	.43	-.33	-	-.27	-	.52		.22	73
44	-	-.30	-.31	-	.51	.24	.41	.44	-	-	-	-	-	.71		
	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	

Appendixes to Part II

A. Forms

B. Instructions to Field Staff

C. Quality Control Procedures

D. Procedures for Hand Analysis of OSCI

E. Interim Report After Site Visit

UCLA FEEDBACK RECORD SHEET

FB Session # ____________________ Date ____________________

Location of Meeting ____________________ Time begun ____________________

Participants: (by name)

Site #	Class #	Head Teacher	Aide	Other (specify)
________	________	________	________	________
________	________	________	________	________
	________	________	________	________

Anticipated Agenda

Rate extent to which points covered

Completely					Not at all	
7	6	5	4	3	2	1

____

____

____

____

Participation by individual teachers (Use reverse side)

Questions asked by teachers (Use reverse side and identify teacher)

Names of specific children about whom teachers wanted information:

List suggestions made by teachers:

UCLA Representatives present at meeting:

Delegate Agency Representatives present:

Report submitted ____________________

Date of report ____________________

UCLA HEAD START EVALUATION AND RESEARCH CENTER

UNIVERSITY OF CALIFORNIA, LOS ANGELES

The Head Start class your child is attending has been chosen as a sample class for the national evaluation program. We are asking permission of the parents in this class to allow their children to take some tests to show us how much the class as a group has learned from the Head Start Program. Most children enjoy "playing the games" and of course there is no danger to their safety or well-being in any of them.

All the evaluations will be carried out by professionally qualified people. Test scores of individual children will be kept completely confidential and no records of your child's tests will be given to anyone without your written consent. In addition, we will be making some photographic records of classroom activities which will be shown to teachers.

Although you do not have to allow your child to be included in the evaluation sample, we are sure you will want to help us obtain information which will lead to building the best Head Start program possible.

You may indicate your permission to include your child in the national evaluation by signing and returning the form below.

. . . .

I give permission for my child ____________________ to take part in
(child's name)
the national evaluation program.

____
(Parent's signature)

____
(Head Start Center)

UCLA HEAD START EVALUATION AND RESEARCH CENTER
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Estamos tratando de hallar una manera o modo a saber si la experiencia de Head Start les ayuda a los niños después que entran al kindergarten. Para hacer esto, necesitamos informarnos de los niños que no han ido a Head Start en comparación con los que han ido. Estamos pidiendo permiso de Uds. para darle a su niño unos examenes.

Todas las pruebas estarán dados por personas profesionalmente calificados. Los resultados de cada niño será completamente confidencial y no daremos los resultados de los examenes de su niño a nadie sin su permiso escrito.

Aunque Ud. no tenga que permitirnos a dar las pruebas a su niño, estamos cierto que Ud. querrá ayudarnos a consequir los informes que nos ayudarán a hacer el programa lo mejor posible.

Ud. puede indicar su permiso que incluyemos su niño en los examenes nacionales por afirmar y volver la forma de abajo.

……

Doy permiso que mi niño ____ pueda participar en los examenes para evaluar la programa nacionale de Head Start.

____
(firma del padre)

____
(Head Start Center)

UNIVERSITY OF CALIFORNIA, LOS ANGELES

HEAD START EVALUATION AND RESEARCH CENTER

TO: ALL EVALUATION FIELD PERSONNEL

FROM: LIZ RAYMER, EVALUATION COORDINATOR

SUBJECT: PROCEDURES TO BE FOLLOWED IN FIELD WORK AT HEAD START SITES

Although there are many different kinds of jobs in the field during the year's evaluation program, there are certain rules and procedures which must be followed by all personnel. If possible, you will be assigned to a limited number of sites and will return to these same sites frequently. In this way you will get to know the teachers and children well. Because you will be returning often to the same sites, it is absolutely imperative that your relations with staff and children remain cordial and warm. Adherence to a few simple rules will help maintain this relationship:

I. Behavior in the Classroom

1. Greeting teachers and children by name is always appreciated - SMILE!

2. You are the GUEST of the site at which you are working. Be friendly but do not disrupt the class schedule by engaging the staff in conversation or bothering them unnecessarily. If there is information you must obtain from staff members, be as brief as possible or make arrangements to meet with teachers at a specified time before or after class is in session.

3. It is essential that you win the confidence of the children with whom you are working. Take enough time alone with them to gain rapport, but never become involved with the children's activities in the classroom unless specifically asked to do so by a teacher. Of course, if a child appears to be in danger you should intervene. Children can learn that while you are observing in the classroom you are doing your work. Socializing must be reserved for free moments.

4. Never get involved in discussions with site personnel about anything controversial. This includes discussions of "negative" behaviors of specific children, giving advice to a teacher about her children or her program, or giving test results to teachers. Any requests of this nature should be referred to the Evaluation Coordinator or the Feedback Analyst.

5. At certain periods during the year Data Aides will find that their work load is less heavy and at these times they should make it known to teachers in their assigned classes that they are available to help out. However, a Data Aide should not direct a class activity and is not permitted to take over the teacher's responsibility.

6. Head Start children should never be transported in University or private vehicles.

7. Never use Head Start site telephones for personal calls. If you have to call the office for any reason, call COLLECT and make it known to site personnel that you have done so. The number is 825-4051.

8. Field staff always carry cards which identify them as personnel of the Head Start Evaluation and Research Center. If there is ever any question about your right to be on a site, present a card or have the questioner call the office.

9. Always be sure the teachers know when you are coming to their sites. We will try to supply schedules giving dates for testing and observing at all classes. Avoid unannounced visits as much as possible. If you know you are returning to the same site the following day, tell the Head Teacher before you leave. The teacher, in turn, will be asked to inform our Center of any trips away from the site. When you call a site to ask whether you can come, do not call during class hours as this disrupts the teacher's schedule. The best time to call is within the first hour after the children have left for home, generally between 1:00 and 2:00 P.M. However, if you have arranged to visit a site and are at the last minute prevented from doing so by illness, car trouble etc., please call the site immediately.

II. Traveling to Head Start sites.

1. Know where you are supposed to be each day and at what time.

2. Be sure you have all your testing or observing materials, kits and forms. If you are traveling to a site from the office, make sure that someone has arranged for transportation.

3. As much as possible, field personnel should travel as a group in University cars. ALWAYS SHOW UP AT THE PREARRANGED TIME AND PLACE. Remember, by being late you waste other people's time as well as your own.

III. Submitting Data.

1. Data will be handed in directly to the data processing department. Procedures for submitting will be reviewed by Frank Wang. Carry completed data around with you as little as possible. The sooner it is in the hands of data processing the less likely it is to be misplaced. If you do not come into the office every day, try to send your data in with someone who does.

2. Be familiar with the basic information which must be filled in on all forms. Although this varies for different instruments, the following identifying information must always be filled in:

Child's name
Birthdate
Site name and number
Class number
Teacher's name
Date test administered
Time of testing
Your name
Your examiner number

Never hesitate to ask questions about the correct way to administer a test, or to gather any kind of data. Errors and omissions are extremely costly and missing data can sometimes never be retrieved. It is smart to ask questions, and not smart to be afraid to ask.

IV. Recording Completion of Specific Tests in Classrooms.

A list of children to be tested will be posted in each sample class in a conspicuous place. Every time an examiner administers a test, it must be entered on this record sheet. There is a space under each required test for the date of administration, the examiner number, and time of test. It is absolutely essential that this information be recorded immediately to avoid missing a child or administering a test to the same child twice.

Order of administration of tests

A. Experimental and Control 1:

Day 1. Stanford Binet with Birch scoring and Factors Affecting Test Performance

Day 2. Gumpgookies

Day 3. Preschool Inventory and Animal House

B. Control 2:

Day 1. Peabody

Day 2. Gumpgookies

Day 3. Preschool Inventory with Factors Affecting Test Performance

V. Removing Children from Classes for Testing.

1. Be sure the teacher is aware when you take a particular child and when you return him to the class. Often she can see what you are doing, but if there is any doubt at all, simply say, "I'm taking John to play games now."

2. Always accompany the child to the testing room and personally return

him to his class. He is your responsibility as long as you have him out of the classroom.

3. Before taking a child out of the class, spend a little time talking to him to gain his confidence. This is most important at the beginning of a testing period. Soon the children will get to know you and they will all want to come with you. Smile! Get yourself physically down to the child's level - stoop, kneel, provide warm assuring physical contact. Never refer to "tests" - these often mean blood tests or something frightening to the child. Ask the child to come and "play a game" with you or "come and see our toys."

4. If a child refuses to take a test, spend a reasonable amount of time attempting to gain his confidence, but never force him to take a test. There are many ways in which children can be helped to overcome fear or shyness. Some of these are listed below.

    a. Take the outgoing children first. They will pass the word or will be observed by the more reticent children and before you know it, they will all want to go with you.

    b. Spend extra time with the hesitant child.

    c. If a child refuses to participate, let him go without making him feel that you are angry with him; the next day greet him with a smile and ask if he is ready for his turn today.

    d. Mexican-American children may feel more comfortable with a Spanish-speaking tester. Although the tests must be given in English, a Spanish-speaking examiner can help gain the child's confidence.

    e. Sometimes a particularly shy child will go with the examiner if a special friend or aide is allowed to go along for company. Although this practice should not be encouraged, sometimes it proves to the child that this is a non-threatening situation. Often the friend or teacher can leave after a short while and the child is then perfectly willing to stay with the examiner.

5. If a child gets extremely restless or refuses to go on with a test, give him an opportunity to stretch, go to the bathroom, or get a drink of water. Examiners should not discontinue a particular test in the middle unless absolutely necessary. Individual sittings have been planned so that they are not too long or tiring for the average child.

6. Find out from the teacher when the children have snack and lunch, and adjust the testing schedule so as not to interfere with these periods.

QUALITY CONTROL PROCEDURES

I. Stanford-Binet and Birch Procedure

1. Only Binet testers who had either completed a course which included administration of the Stanford-Binet scale to several subjects or had given the test to preschoolers, were appointed.
2. One of these examiners, a doctoral candidate in the Pupil Personnel area of Educational Psychology, spent several weeks studying the Birch scoring system and practicing on preschool subjects. All the practice sessions were videotaped for review and criticism. After proficiency had been achieved, this person was assigned the task of training all the Binet testers in the use of the Birch procedure.
3. Every Binet tester received six-to-eight hours of training. This consisted of discussion, practice, and final administration under supervision of the Binet-Birch measures to at least one four-year-old child.
4. All protocols were checked by the trainer; the examinations given by the trainer were checked by one of the qualified examiners.

II. Gumpgookies, PSI, WPPSI Animal House

1. Testers were required to become familiar with the test materials.
2. Testers practiced administering tests to each other. Each trainee was videotaped for 10-15 minutes during the practice test sessions. These tapes were then played back and discussed.
3. Those trainees who required further training made tape recordings as they administered additional practice tests; they then played back the tapes while checking with the manual to detect errors.
4. The first few regular protocols were carefully checked by the Evaluation Coordinator. All test protocols were scored by one staff member in addition to the examiner.

III. OSCI

1. There was an extensive training program for this instrument at UCLA since there were many problems which had to be worked out before the national training sessions could be scheduled. Of the eight UCLA trainers, three of the most proficient were selected to serve on the national training team with the Head Trainer.
2. Every observer-trainee reached prescribed reliability with the Head Trainer during a series of practice observations. Those selected as trainers were also required to reach reliability with each other.
3. Before every series of observations, every observer was given one or more practice sessions with the trainer to check the reliability.
4. Observers were required to take a written test prior to each observation series, and had to obtain the same reliability score as required on a national basis. From time to time, reliability checks,

involving two observers focusing on the same activity, were carried out at the UCLA sites. In addition, simultaneous observations with the Group OSCI, the Individual OSCI, and anecdotal reports, were made to provide data for validity checks. The analysis of these validity tests will be included in the OSCI analysis.

IV. Post Observation Teacher Rating Scale (P.O.T.)

1. An introductory meeting was held and the rationale of the instrument and the scoring of the items were discussed.
2. Because of the ambiguity of many of the items, there were extensive discussions with the testers in an attempt to reach some degree of consensus.
3. Where consensus could not be obtained, questions were referred to the Syracuse Head Start Evaluation and Research Center, where the instrument had been drafted.
4. The first run of the P.O.T. was completed by pairs of observers who spent a whole morning in the classroom and attempted to find some agreement in rating the items.
5. The post-OSCI scoring of P.O.T. was done jointly by both members of the observer team. When two OSCI observers were not required for the observation, the one who was not doing the OSCI coding was instructed to attend to the teacher behavior relevant to P.O.T. items.

V. Parent Interview

1. All the staff who were to administer the Parent Interview participated in a two-and-a-half-day training program conducted by representatives of the Washington office.
2. Every completed schedule was checked as soon as possible for consistency and for missing data by an independent interviewer.
3. A list of interviewer errors and omissions were kept on file and immediate feedback was given to the appropriate examiner. Most of the omissions resulted from insufficient probing to get responses which could be coded after the interview.

VI. Descriptive Data

The Data Aides assigned to the specific sites were responsible for keeping the attendance records and obtaining the data for the Child Master Data Form, Staff Characteristics, Classroom Facilities, and Administrative Variables. Other Evaluation staff familiar with a particular site made occasional spot checks to maintain consistency and accuracy.

VII, Census Forms and Health Records

The Evaluation Coordinator met with the teacher in each agency to arrange for completion of the various census forms. The Medical/Dental form was filled out by the nurse. In addition, arrangements were made to obtain height, weight, and blood analysis data through the regular medical facilities of the particular agency.

PROCEDURE FOR ANALYSIS OF DATA FROM CLASSROOM OBSERVATIONS

Based on the 1967-1968 experience, it seemed quite unlikely that the results of the computer analysis of the 1968-1969 OSCI would be available in time to meet the deadlines for submission of Final Reports by the various Evaluation and Research Centers. For this reason, a technique was devised to enable the Centers to do a quick analysis which they could incorporate in their own reports. Information such as the percentage of time spent indoors and outdoors, the variety of materials used, classroom variables, and teacher data can be obtained by following the procedures indicated.

I. General Information

1. To determine the percentage of time spent indoors and outdoors:

   a. Tally the number of observation units spent indoors and the number of units spent outdoors. Where both are circled, tally half as indoor and half as outdoor units.

   b. Divide the total number of tallies in each group by 40, which is the total number of units in an observation day.

2. Materials:

   a. Make a tally of the materials recorded in an observation day.

   b. Make a list of all the different materials used.

   c. If a type-token ratio is desired, divide the total number of entries (a) by the number of different types of materials (b).

II. Class Observation Data

1. To obtain average size of child groupings:

   a. Tally number of records used in class observation sections.

   b. Use the following designations for group sizes: 1 = one child; 2 = 2-5 children; 3 = 6-10 children; 4 = 11 or more children. Divide number of groups in each category of b by total number of records in a.

To obtain percentage of

2. Children in child-controlled activities and those in non-child controlled activities:

   a. Tally number of children in each class observation unit over the whole observation day. Use the actual number of children observed, not the totals at the bottom of each class observation unit.

b. Tally the number of children in activities where Locus of Control is coded "1," indicating pupil is in control.

c. Obtain number of children in activities not under child control by subtraction (a - b).

d. To obtain percentages of children under child vs. non-child control, divide both b and c by a.

3. Time spent in routines:

a. Tally number of children in records having Context Codes E, I, C, R, or T when used with Content Codes other than la, ss, sc, qu, vd, ad, vm, or po.

b. Divide 3 a by total[1] number of children.

4. Time in which children are uninvolved or doing nothing:

a. Tally number of children in records with U na and W na.

b. Divide 4 a by total number of children.

5. Time spent in Cognitive activities:

a. Tally number of children in records coded la, ss, sc, or qu.

b. Divide 5 a by total number of children.

6. Time spent in Music, Art, Dance, Dramatic activities:

a. Tally number of children in records coded ar, mu, da, or dr.

b. Divide 6 a by total number of children.

7. Time spent in small muscle, building, and sensory activities:

a. Tally number of children in records coded S (except S dr), B (except B with dr, la, ss, sc, or qu.), and W (except W with la, ss, sc, or qu.)

b. Divide 7 a by total number of children.

8. Time spent in large muscle activities:

a. Tally number of children in records coded L (except L with dr, la, ss, sc, or qu.)

b. Divide 8 a by total number of children.

---

[1] In all subsequent computation, obtain total number of children as in II 2 a.

III. Teacher Observation Data

1. Teacher Involvement: The total number of 20 second teacher observation records over the entire day is 144. This includes 72 T_1 records and 72 T_x records.

a. Tally involvement coded as 1, 2, or 3 separately, and 4 + 5 together, for all T_1 observations.

b. Tally as above for all T_x observations.

c. Divide totals under a and under b by 72.

d. Totals over both Head Teacher and Assistant can be obtained by summing a and b and dividing by 144.

2. Size of Groups with which Teachers are Involved:

a. Assign groups to categories according to following code:
1 = one child, 2 = 2-5 children, 3 = 6-10 children,
4 = 11 or more children, 5 = no children.

b. Tally number of groups in each category over entire day.

c. Divide each of groupings under b by 144.

3. Teacher Input:

a. Tally input content codes for each T_1 and T_x set of records, according to the following categories:

1) sv, sp, si, vc (socializing and verbal communication);

2) la, ss, sc, qu (cognitive); 3) vd, ad, vm, po (sensory);

4) sk (skill); 5) ru (rules); 6) No input.

No more than one entry per category is made for each record. For example,

$\boxed{vc}\boxed{sv}\boxed{x}$ receives one tally in category 1.

$\boxed{la}\boxed{qu}\boxed{vd}$ receives one tally in category 2, and one tally in category 3.

b. Divide totals within each category by 72 for T_1 and 72 for T_x.

c. Totals over both Head Teacher and Assistant combined can be obtained by summing T_1 and T_x entries for each category and dividing by 144.

INTERIM REPORT AFTER SITE VISIT

STUDY IDENTIFICATION

Evaluation and Research Center:	University of California, Los Angeles
Center Director:	Carolyn Stern
Evaluation Coordinator:	Elizabeth Raymer
Major Purpose of Study:	Effect of Evaluation Feedback
Person Visiting Site:	Dr. Edward Suchman
Date of Visit:	February 4, 1969

STUDY CHECKLIST

1. Overall Design: No change

2. Hypotheses:

There has been no change in the hypotheses as stated in the original proposal. However, it now seems that this feedback intervention study may be serving a much more important training function than had been expected. Its heuristic value depends not only upon what the teachers do with the information given them, but also upon how involved individual agency supervisory staffs have become in the feedback.

A new hypotheses might be stated as follows: The closer the involvement of child development supervisors and agency trainers, the stronger the effects of the feedback program on the children, the teachers, and the parents. Since records are being kept of differences in intensity of involvement by the four participating agencies, this should be a testable hypothesis.

3. Sample:

There have been no changes in the size and type of sample planned. None of these classes were observed before the allocation to treatment was made. The Director of Head Start from EYOA arranged a meeting of the 12 Delegate Agency representatives. At this meeting the design for the Evaluation Intervention was presented (see page 24) and participants were invited. Nine of the 12 agencies were interested and selections were made so as to include agencies which had not been involved in the previous evaluations (1966-67 and 1967-68). Additional criteria were (1) eagerness to participate and (2) inclusion of a variety of environments.

Classes were selected from one small Community Action Program and from three Delegate Agencies under the Los Angeles Economic Youth Opportunities Agency. The classes from each agency were randomly assigned to the three treatment groups, with the restriction that all classes within a center were to be assigned to the same treatment.

ASSIGNMENT TO TREATMENTS BY AGENCIES

Agency	Number of Classes by Treatment			
	Feedback	Control 1	Control 2	Total
I	4	2	-	6
II	2	2	2	6
III	2	3	2	7
IV	2	1	2	5
	10	8	6	24

4. Variables, Measures, Instruments

a. Independent and background variables: Amount of intervention or feedback is the main independent variable. The background variables to be considered are teacher, class, child, program, and agency characteristics.

b. Measures and instruments: In addition to the national evaluation package, two instruments, the TATE and the TEACH, have been developed to measure teachers' attitudes towards evaluation and teachers' educational objectives. Codebooks have been developed and some analyses have been run on the pilot testing of these instruments. The TEACH has been given to 36 additional adults so as to obtain a broader data base on this instrument.

A series of tasks are being developed to measure children's achievement on those instructional goals considered important according to the responses on the TEACH scale. A selection of these tasks will be administered to sample children during the posttest period.

5. Testers and Observers

Training of testers and observers was completed in September and October of 1968. Procedures for training and quality control statements have been sent to the Washington office. A copy of these guidelines is included on pages 118 to 119 of this report. Turnover has not been a problem nor is it expected to be. Only one member of the observation staff resigned, but there are enough testers for posttesting without additional recruitment. There will be a retraining period prior to posttesting.

6. Data Collection

a. Status of Data Collection: All pre-data for the national evaluation have been collected and the tape will be sent to Washington within the next week or ten days. There were only three untestable children, and these were not untestable on every measure. Data collection did not conform precisely to time guidelines, but it has gone far more smoothly compared to previous years, particularly considering that the individual child testing is much heavier than ever before.

b. Problems: The most severe problems were not unanticipated. The main difficulty was absence of sample children due to colds and other respiratory infections associated with an unusually rainy winter. Since schools in Southern California are geared for a great deal of outdoor activity, inclement weather always results in irregular attendance.

There has been some degree of resistance to the evaluation program expressed at several centers and anxiety at all of them. There were verbal expressions of hostility from teachers at the orientation meetings. Some teachers felt that they had been put into the evaluation program by their supervisors without being consulted. Several teachers felt that the individual testing would be disruptive for the children. This is probably a rationalization on the part of the teachers, who are projecting their own fears onto the children. After the orientation meetings, most of the resistance was dissipated; one teacher in the feedback group challenged the validity of the measures and so a special meeting was arranged to answer her questions. At one control site the teachers have been passively uncooperative and unfriendly, and their attitude has not changed.

At two sites there were objections to some sections of the Parent Interview. This was handled by arranging to have the Evaluation Coordinator meet with parent groups to explain the purpose of the interview and stressed the right of parents to refuse to answer any questions which they found objectionable.

Another not unanticipated problem concerns teacher attrition. So far six of the 24 original teachers have resigned or been transferred. Fortunately, this attrition is spread evenly across all treatments. Other problems were less severe than expected. There was some initial resistance to the program on the part of certain teachers, but these problems have now been ironed out.

c. Analysis: Data is being processed as obtained. As soon as the pre-tape is completed, coding and punching for the mid-year measures will begin. The data from the six classes in Control 2, which receive the minimal evaluation package, will also be analyzed.

To date, only the factor analyses on the TEACH and the TATE have been run.

The major hypothesis will be tested by analysis of variance for pre-post changes of teacher responses on the TEACH and TATE scales. These will be related to changes in teacher classroom performance as measured by the OSCI and to changes in children as measured by the national evaluation instruments.

7. Treatments

The monthly feedback meetings have a general form which is outlined in the intervention proposal. Plans are firmed up before each series of meetings and the agenda for all feedback meetings is kept standard within treatments.

Some departures from the original proposal have been made, but these are primarily in terms of scheduling of meetings to accomodate minor changes required to dovetail results of tests and videotaping with scheduled meetings.

Treatments are being applied on a regular monthly basis as planned, except that the control groups attended only one Orientation Meeting. Treatment implementations are being monitored in several ways. The meetings are tape recorded and reviewed by the Evaluation Coordinator, who has attended several feedback sessions. The written reports of each meeting and the tapes are available to the Coordinator and the Director, as well as to any other interested investigators.

8. Schedule

An overview of the schedule is included on pages 38 and 39 . The activities have been proceeding according to this plan.

9. Coordination

No coordination with other Head Start Education and Research Centers was planned. The Boston University Head Start Evaluation and Research Center is carrying on a study which has some similarities with the UCLA intervention. The Director of the Boston Center visited Los Angeles and spent some time at a Feedback Meeting and in a subsequent discussion with the UCLA Evaluation staff.

10. There are no other comments at this time.

PART III

OSCI Training Report

OBSERVATION OF SUBSTANTIVE CURRICULAR INTERACTION[1]

Training Program for National Head Start Evaluation

1968-1969

As part of the national evaluation program, all 12 Evaluation and Research Centers were required to carry out a series of curricular observations in each of the sample classes. The instrument used was the Observation of Substantive Curricular Interaction (OSCI). Since the OSCI had been developed at UCLA, the responsibility for training the observers at each of the E & R Centers and for maintaining quality control was delegated to UCLA. This task was assigned to Arrie Bachrach, OSCI Training Coordinator of the UCLA Center staff.

Each E & R Center was responsible for four cycles of OSCI observations, originally scheduled for November, January, March, and April. Since several Centers were unable to get under way, due to various local problems, this schedule was later modified; however, the training was arranged so that the first round would be completed by the beginning of November.

The number of trainees (being dependent on the personnel available as well as the size of the evaluation sample of a particular E & R Center) ranged from two at South Carolina to eight at Temple. A total of 65 potential observers were trained. Because of the close supervision and one-to-one instruction required for this training, two qualified UCLA trainers were sent to those locations which had more than three trainees. To encompass all the training within a limited time, it was necessary to conduct training sessions simultaneously at several E & R Centers. Additional personnel were therefore selected from the most competent observers trained at UCLA. These trainers had achieved .86 to .88 reliability with the OSCI training coordinator.

PROCEDURE

The order of training sessions took into consideration both the needs of individual E & R Centers and their geographic locations. Thus, as far as possible, the schedule was set up so that Centers located in the same part of the country were assigned contiguous training dates. The final training schedule is presented in Appendix A.

Once acceptable dates had been determined, each Center was instructed to make the necessary arrangements for the training session, including accommodations for the trainers and appropriate sites for field practice. Manuals, sample coding forms, and practice coding exercises were prepared and mailed out well in advance of the scheduled date and the trainees were instructed to memorize the codes, work the coding exercises, and attempt one or two practice observations.

---

[1] The Observation of Substantive Curricular Interaction was developed at the University of California Head Start Evaluation and Research Center, under the direction of Dr. Carolyn Stern, OEO Project #4117. The original version was used in the 1967-68 National Head Start Evaluation; the present modified version was used in the 1968-1969 evaluation.

After the team of OSCI trainers arrived at a Center, the training followed the standard sequence described in Appendix B. In those cases where trainees did not achieve the required reliability before the UCLA trainers left, the local observer who had achieved the highest reliability was delegated to continue working with the local staff until the required level of reliability (.70) was achieved. This reliability was calculated by comparing the trainer's and trainee's records of the same classroom situation and computing the percent of agreement between the two records. (See Appendix C for method of calculating observer reliability.) Observers also had to demonstrate written mastery of the OSCI coding system by passing, with a minimum score of 70%, a written test administered by the OSCI trainers at the conclusion of the training session. (See Appendixes D - G.)

Quality control during the year was maintained by sending out a series of coding tests, together with answer keys, so that the local evaluation coordinator could administer the test and provide immediate corrective feedback to the local observers. The tests themselves were sent to UCLA and corrected by the OSCI training coordinator. Test scores together with comments were then returned to the various E & R Centers. Observers who had failed to achieve the criterion of 80% correct were required to take remedial training before being permitted to embark on the next observation cycle.

RESULTS AND DISCUSSION

Table 1 presents the means and standard deviations for reliability and posttest scores by E & R Centers. In all cases, the posttests administered by the local evaluation coordinators show higher scores than those administered by the UCLA trainers. This may be due to the fact that the number of people taking the tests decreased, with hopefully the least competent trainees either dropped or shifted to other tasks, and those remaining improving with practice. Unfortunately, it is also quite likely that the local coordinator may not have been completely impartial in the test administration.

An important variable in determining whether trainees would achieve the reliability criterion was the extent to which the pretraining exercises had been carried out. At Kansas, Texas, and Hawaii, where the trainees had memorized the codes, completed the coding exercises, and attempted to use the OSCI in the field, the subsequent training sessions went smoothly and all abservers reached criterion. However, at several Centers the personnel, who had been employed a day or two before the training session began, had little or no knowledge of evaluation in general let alone familiarity with the OSCI codes and rationale. These trainees could not be expected to reach field reliability criterion during the three-day session. Unfortunately, the UCLA trainers had neither budget nor time (because of the tight schedule) to remain more than the allotted period. The compromise solution was to bring at least one local ob server to the reliability criterion and then have this person assume responsibility for the trainees at that E & R Center.

Training sessions also went smoother if any of the trainees had been OSCI observers in 1967 - 1968. Although these observers experienced some

Table 1

E & R CENTERS	Inter-rater Reliability September N	M	S.D.	Posttest Scores October N	M	S.D.	January N	M	S.D.	March N	M	S.D.
UCLA[a]	9 4	85.4 92.3	4.45 2.28		b		7	92.3	2.00	7	92.4	3.46
TEXAS	4	85.3	1.48		b		4	94.0	0.23	3	93.0	2.45
SOUTHERN [c,d]	8	70.0			b		7	91.7	3.20	7	96.3	1.23
TULANE	4	80.3	3.10		b		4	90.0	1.58	4	87.0	2.55
MICHIGAN	5	80.4	3.93	5	94.0	2.29	4	91.8	2.00	4	91.0	2.75
KANSAS	5	79.2	3.37	5	91.4	5.92	4	88.8	2.58	4	90.5	1.12
BANK STREET	4	86.5	4.39	4	95.3	2.05	3	89.3	4.78	3	87.0	4.32
BOSTON	4	85.0	3.24		c			c			c	
SYRACUSE	8	83.5	2.83	8	88.6	4.11	6	90.5	4.07	3	86.7	2.26
TEMPLE	7	84.3	7.91	7	89.1	6.81	7	91.4	3.97	7	88.0	4.81
SOUTH CAROLINA	2	92.0	1.00	2	94.0	2.00	2	92.0	2.00		c	
HAWAII	6	89.0	4.47	6	88.5	2.22	5	94.0	1.90	3	92.0	0.81

---

[a] UCLA had a second reliability check in April.

[b] These were the first four centers trained; they had no October posttest.

[c] Center did not return data for this measure.

[d] Only one of the trainees at Southern achieved the minimum 70.0% Inter-Observer agreement.

interference due to the changes in the new form of the OSCI, the general state of preparation of the whole training group at any Center was much better where the trainees had had the benefit of an experienced OSCI observer among them during pretraining.

In general, the Evaluation and Research Centers were cooperative with the quality control procedures and returned the test protocols to the UCLA OSCI coordinator. The exceptions were South Carolina, which returned only one set of posttests, and Boston, which returned none and failed to respond to any of the letters requesting this material.

SUMMARY

During the 1968-1969 National Head Start Evaluation, the Head Start E & R Center at UCLA conducted three-day OSCI training sessions at each of the E & R Centers. Between September 10, 1968 and November 8, 1968, 65 observers were trained, with 55 of these trainees achieving reliability during the training sessions. All observers who returned the OSCI Coding Tests received passing scores of 80% or above.

The UCLA staff developed training materials, administered the training sessions, maintained quality control throughout the year by a series of coding tests, and responded to a variety of inquiries and questions from the network of E & R Centers. In addition, the UCLA Center supplied all the printed materials and forms both for the training and for the evaluation sample.

Appendixes to Part III

A. Schedule of 1968 OSCI Training Sessions

B. OSCI Training Procedures Used by UCLA OSCI Trainers

C. Calculation of Inter-Rater Observer Reliability

D. OSCI Training Exercise

E. OSCI Coding Posttest

F. Coding Test for OSCI Observers #1

G. Coding Test for OSCI Observers #2

APPENDIX A

Schedule of 1968 OSCI Training Sessions

Training of UCLA Observers, September 5-15, 1969
by Arrie Bachrach, Training Coordinator

Trainees: Ruth Silberstein
Karen Stingle
Sandy Mitchell
Ila Baum
Eva Benesch
Sharon Nichols
Elma Forrest
Judy Blum
Paula Board

Trainers: Arrie Bachrach and Ruth Silberstein

University of Texas, September 25-27, 1969

Trainees: Nancy Templeton Cross (Evaluation Coordinator)
Thad Eckman
Rick Reposa
Ann Maurer

Southern University, September 30-October 3, 1969

Trainees: Johnnie Smith
Marge Jackson
Helen Burton
Lillie Louis
Velma Wilkerson
Georgiana Johnson
Wanda Williams
Amelia Wilkerson

Tulane University, October 2-4, 1969

Trainees: Corinne Duffy (Evaluation Coordinator)
Juanita Quimby
Virginia Sicotte
Linda Nathey

Trainers: Karen Stingle and Sandy Mitchell

Michigan State University, October 14-16, 1969

Trainees: Donald Meltzer (Evaluation Coordinator)
Jo Lynn Cunningham
Chris Walter
Harriet Solomon
Eileen Earhart
Linda Smith

Kansas University, October 16-18, 1969

Trainees: C. Wallis Henning (Evaluation Coordinator)
Betsy Phillips
Rob Robinson
Brad Kahler
Cori Brown
Joan Blasdel

Trainers: Ruth Silberstein and Karen Stingle

Bank Street, October 21-23, 1969

Trainees: Doris Wallace (Evaluation Coordinator)
Shirley Smith
Janice Roth
Rebecca Wallace
Kate Oram

Boston University, October 23-25, 1969

Trainees: Bunny
Marge
Nancy
Flo

Syracuse University, October 28-30, 1969

Trainees: David Lindstrom (Evaluation Coordinator)
Nancy Brisson
Rosemary Trubitt
Sherry Schultz
Joan Manzelman
Linda Hollingshead
Dolores Peterson
Dorothy Mullen

Temple University, October 31- November 4, 1969

Trainees: Judy Grodwin (Evaluation Coordinator)
Peggy Howard
Charles Slater
Nancy Weidenfield
Natalie Sanders
Pat Newman
Pat Phillips

Trainer: Karen Stingle

University of South Carolina, October 31-November 1, 1969

Trainees: Sharon Pearce
Carolyn Corbitt

Trainer: Arrie Bachrach

University of Hawaii, November 6-8, 1969

Trainees: Hannah Herman (Evaluation Coordinator)
Patsy Hart
Carole Hodges
Betty Elrod
Annette Okimoto
Ginny Lerner

APPENDIX B

OSCI Training Procedures Used by UCLA OSCI Trainers

Fall 1969

Day 1:

AM 1. Brief period for answering questions.

2. Administer Pretest on Coding.

3. Present material from Manual.

4. Review coding principles.

PM 1. Field practice session I.

2. Return to office to discuss field practice.

Day 2:

AM 1. Field practice session II.

2. Return to office to discuss field practice.

PM 1. Correct and discuss OSCI coding exercises.

2. Discuss OSCI rationale so as to elucidate coding decisions.

Day 3:

AM 1. Field practice session III. Reliability check.

2. Return to office to discuss reliability computation.

PM 1. Final discussion; question and answer period.

2. Administer written posttest. (See Appendix E.)

3. Review posttest responses.

APPENDIX C

Calculation of Inter-Rater Observer Reliability

Reliability is calculated on the basis of records made by two observers focusing on the same situation. For the Class observations, each entry within a box is given one point, with a possible total of seven points if there is complete agreement between the two records. One point is given for agreement on each of the following: 1) locus of control; 2) context; 3) first content code; 4) second content code; 5) social interaction; 6) number of children in group; and 7) materials.

For the Teacher Observation, eight points are possible. There is one point for each of the three teacher input boxes, two points for degree of teacher involvement, and two points for description of child activity during teacher observation. For the latter, it is possible to earn one point partial credit if there is agreement on both context and content code, even if there is disagreement on additional content codes and materials. However, no partial credit is given for agreement on context and materials or content and materials alone.

A. Computing Reliability of Classroom Observation:

Observer A					Observer B					Agreement Points	Computation
INDOOR C OUTDOOR					INDOOR C OUTDOOR						
1	X				1	X					
1	me	X	X	4	1	na	X	X	4	6	39 Points 42 Possible
1	Swings				1	Swings					
L	me	X	X	2	L	me	X	si	2	6	Reliability Index for Classroom Observation = .93
1	Bike				1	Bike					
L	me	X	X	2	L	me	X	X	2	7	
1	X				1	X					
W	na	X	X	1	W	na	X	X	1	7	
1	Sandbox				1	Sandbox					
S	me	X	X	1	S	me	X	X	1	7	
1	Bike				1	Bike					
L	me	X	sv	2	L	dr	X	sv	2	6	

B. Computing Reliability of Teacher Observation:

Observer A					Observer B					Agreement Points
INDOOR C OUTDOOR					INDOOR C OUTDOOR					
20	Dsc				20	Dsc				
1	sc	sv	X	2	1	sc	X	X	2	7
40	"				40	"				
1	sc	X	X	4	1	sc	X	X	4	8
60	"				60	"				
1	sc	sv	X	4	1	sc	sv	X	4	8
20	"				20	"				
1	sc	X	X	4	1	sc	sv	X	4	7
40	"				40	Nna si				
1	sc	si	X	1	1	si	X	X	1	5
60	Nna si				60	"				
1	si	X	X	1	1	si	X	X	1	8
20	Svm		puzzle		20	Svm		puzzle		
2	vc	X	X	1	1	vc	X	X	1	6
40	"				40	"				
1	sv	X	tu	1	1	sv	X	X	1	7
60	"				60	"				
1	sv	vc	X	1	1	sv	vc	X	1	8

Computation

$\frac{\text{64 Points}}{\text{72 Possible}}$

Reliability Index for Teacher Observation = .89

Overall Observer Reliability

$$\frac{39 + 64}{42 + 72} = \frac{103}{114}$$

Total Observer Reliability Index = .90

APPENDIX D

OSCI Training Exercise. Part I.

Code items 1-22 as both Class and Teacher Observations, using the OSCI protocol form. Number each record to correspond with the test items.

1. The Head Teacher is reading a story to four children. There is no elaboration or participation by the children.
2. Assistant Teacher is sitting with four children, listening to a record. She reminds several of the children to be quiet and listen to the music.
3. Two children are building with floor blocks. They are talking to each other about what they are building. The Head Teacher comes over and watches the children without saying a word.
4. Two children are already painting and the Assistant Teacher is helping two other children put on smocks. As she helps them, she tells them what to do next.
5. Assistant Teacher puts a large tub of water on a table. She turns to four children watching her and suggests that they might like to sail boats in the tub.
6. Child stands looking at cleanup activities. Head Teacher tells him that everyone must put away his share when it is clean up time.
7. It's snack time. Six children are sitting around a table chatting while the Head Teacher is in the kitchen area pouring juice. One child is handing out napkins.
8. Assistant Teacher is reading a story about boats and harbors to six children.
9. In the next class observation unit, the Teacher and the six children are having a discussion about the story.
10. Head Teacher, holding rabbit on her lap, is describing its eating habits to two children. The children are petting the rabbit while the Teacher talks.
11. Assistant Teacher is straightening up doll corner. She looks up and one child who is playing with puzzles smiles at her. She says hello and goes back to work.

12. Whole class is watching a movie about animals in the jungle. Several children are giggling. The Head Teacher tells them to be quiet and watch the movie.

13. Assistant Teacher is reading a book about policemen, firemen, and farmers, and showing the children the pictures. Nine children are watching and listening. One child is on the other side of the room looking at a book.

14. Assistant Teacher is preparing the table for painting. Three children are helping her. She's directing them; telling them where to put materials and praising them for being good helpers.

15. Head Teacher is sitting with three children, presenting an instructional program on numbers. She is holding up cards and asking individual children to identify the numerals.

16. Four children are in the sandbox digging with shovels. They are chattering with each other. The Head Teacher comes over and tells the group that it's going to be clean up time soon.

17. Head Teacher is playing the guitar and singing songs with the whole class.

18. Whole class is resting quietly on cots before lunch. Assistant Teacher is quietly supervising.

19. In the next class observation unit, the class is still resting, but the teacher puts on a record of quiet music.

20. Head Teacher is sitting down watching the class. There are three children riding bikes, three playing trains. One child is on the swings, another child is pushing him, and the two are laughing and talking together.

21. Head Teacher is helping four children make collages. She reminds them not to use too much glue, etc. Another boy is making a tower out of interlocking table blocks. He calls to her; she turns, smiles, and says, "What a nice tower, Bobby." She then continues helping the other children with their collages.

22. Head Teacher playing language lotto with four children.

For the following items, code Class Observations only:

23. One child is sitting on the rug silently watching several children playing with blocks at the other end of the rug.

24. Three children are playing with clay. One of the children is squeezing the clay tightly in his hands so that it runs out between his fingers. He is watching it and laughing.

25. Two children are on the see-saw, singing "See saw, Margery Daw."

26. A child gets up from the puzzle table, runs outdoors and looks around for something to do.

27. In the next class observation unit, the same child is still standing outside, watching two children riding bicycles.

28. In the next class observation unit, the same child is wandering around aimlessly.

29. Child leaves an activity and goes to the bathroom.

30. Two children are sitting in the sandbox, laughing and talking, pushing, each trying to tip the other one over.

31. A boy and a girl are in the doll corner. The girl is holding a doll, the boy is wearing a stethoscope and listening to the doll's heart.

32. One child is hitting a punching bag.

33. A child has just finished painting at the easel and is walking over to the sink to wash his hands.

34. Two children are listening to records. Child 1 picks up the arm of the record player, Child 2 says "Don't do that," and hits Child 1. Child 1 starts to cry.

35. Two children are sitting at a table in the housekeeping corner baking clay cookies.

36. One child is sitting at a table playing with table blocks.

37. Two children pick up drums and begin marching around the room banging a rhythm.

38. In the next class observation unit, a volunteer comes over and says, "Move the way the rhythm makes you feel."

OSCI Training Exercise: Part II

Code the following as a 3-minute Teacher Observation in successive 20 second units. Use appropriate column of OSCI protocol.

1. Head Teacher sits down at a table where three children are playing with puzzles. She watches quietly, not saying a word.

2. She continues watching. Towards the end of the 20 seconds she says, "You're doing very well" and gets up and leaves.

3. Teacher walks around the room making appropriate but casual comments to various children. She praises a group of children; then reminds a girl to put away her materials when she's finished with them.

4. Teacher walks into the kitchen and starts gossiping with the cook. Remains there for this entire 20 second segment.

5. Teacher still talking with the cook. She notices one child climbing on the table and calls out, "Get off there! We don't climb on the furniture in school, Jimmy!"then she resumes talk with cook.

6. Teacher walks back into the classroom, sits down and watches the entire class. She calls out to the class, "It's going to be clean up time soon, so finish whatever you're doing."

7. She sees two children arguing and shoving each other over use of a puzzle. She goes over and puts her arms around both of them, calms them down and says, "You have to share things in school."

8. She resumes walking around the room. A child asks her for help in using the scissors. She stops, shows the child how to use the scissors, and gets involved in a discussion about what the child is making.

9. Teacher gets up and starts to clean up things scattered around the room so that lunch can be served soon.

APPENDIX E

OSCI Coding Posttest

1. The whole class is on the playground and the head teacher is lining them up to go inside. She says, "It's time to line up," and gently shoves a couple of children into place. The children are laughing and poking at each other.

2. Four children are sitting at a table. One of them is building a tower with blocks, and three are putting puzzles together. The teacher is standing nearby watching, and at one point she says to the one child, "What a nice tower you're building."

3. Three children are painting at a table. The head teacher is asking them to name the colors they want before she gives them out.

4. The head teacher is bringing out cots before rest period. While she is setting them up, she sees two boys hitting each other and laughing. She says, "I've told you before that if you want to roughhouse, you have to wait until we go outside." She then returns to setting up the cots.

5. The assistant teacher is reading a book about stores and storekeepers to five children. She sees another child wandering around aimlessly and tells him to come join the group.

6. Head teacher is acting out a rhyme and eleven children are saying it along with her and imitating her movements. Two girls are holding hands and looking at the group.

7. The teacher is on the playground talking to another teacher. One boy is pulling a wagon with another boy in it. Three girls are standing together: two of them are discussing the third, their "patient;" one is using a stethoscope and the other is holding the "patient's" hand and telling her, "It doesn't hurt."

8. All fifteen children are resting and the head teacher has put on a record of quiet music. At the start of the next 20-second interval the teacher begins singing to them along with the record.

APPENDIX F

CODING TEST FOR OSCI OBSERVERS

Prepared by

UCLA Head Start Evaluation & Research Center

November 1968

Part I

E & R Center____

Observer________________________________ Date_______________

Part II

Write the appropriate Context Code, Content Code, Locus of Control Code, or Teacher Involvement Rating to cover each activity or input as it is read to you.

A - Context

1________
2________
3________
4________
5________
6________
7________
8________
9________
10________

B - Content

11________
12________
13________
14________
15________
16________
17________
18________
19________
20________

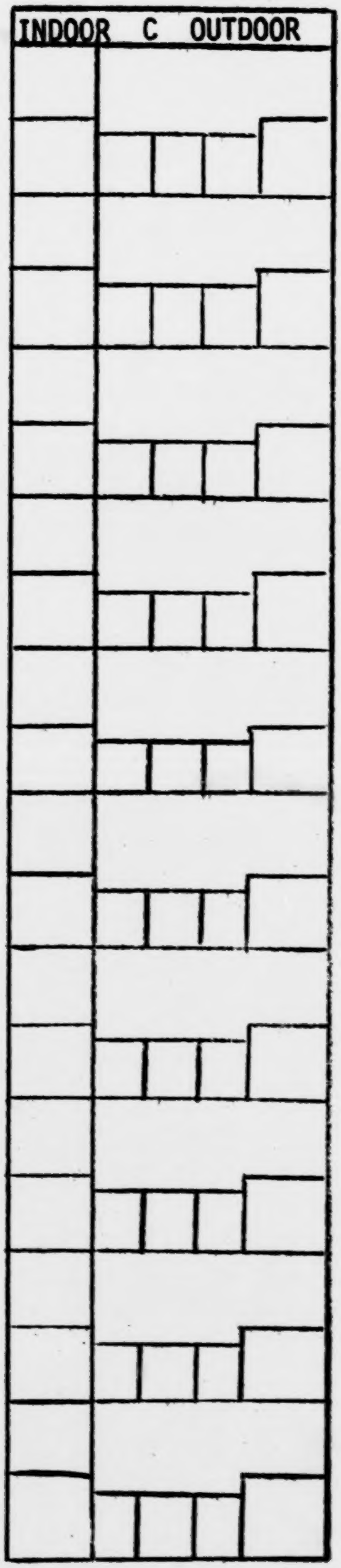
21________
22________
23________
24________
25________

C - Locus of Control and Teacher Involvement

26________
27________
28________
29________
30________
31________
32________

Part III

At a signal from the test administrator, begin coding the class observation unit. When you finish, go on to the teacher observation unit.



Code For Class Observation

Five children and the teacher aide are making collages at a table. The aide is complimenting the children on how pretty the collages are turning out.

Two children are building a farm house out of wooden blocks. They are discussing how many cows they are going to have on their "farm."

Three little girls are in the housekeeping corner wearing dress-up clothes.

One child is staring out the window at nothing in particular.

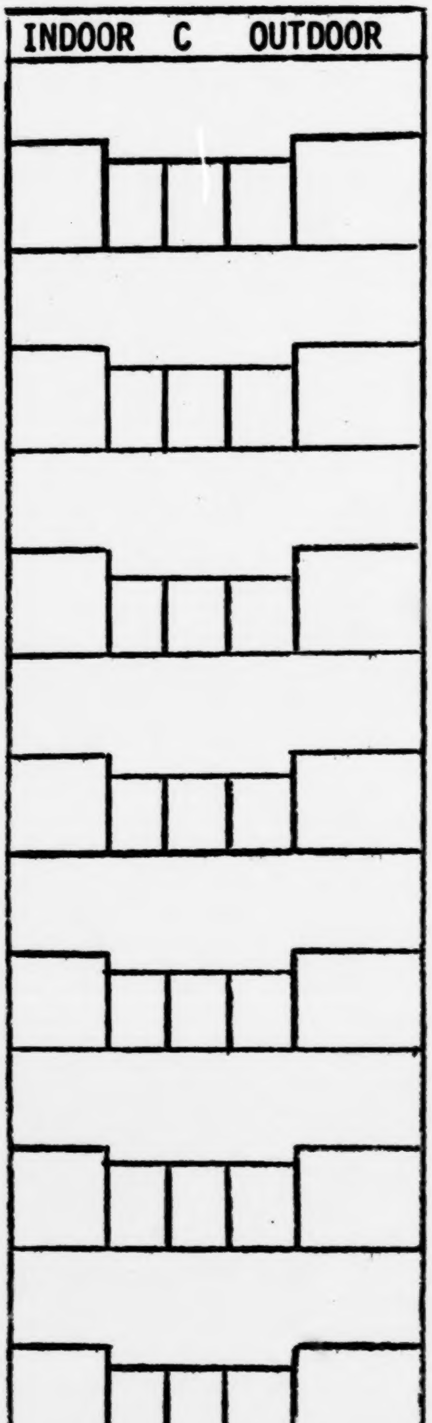
One child is looking at a book by himself.

One child is in the bathroom by himself.

The Head Teacher is helping one child tie his shoe laces and at the same time is showing him how to do it.

One child is working a puzzle.

Code for Teacher Observation



20 Seconds:
Head Teacher is presenting a reading readiness program. She is holding up flash cards and having the children identify the object pictured. There are ten children.

40 Seconds:
The reading readiness program continues. Several times during the twenty seconds she praises children for correctly identifying the objects on the cards.

60 Seconds:
The reading readiness program continues. Towards the end of the twenty seconds, the teacher tells the whole class (including the group she is with), "It's going to be lunch time soon, so why don't we all go wash up for lunch while I get things ready."

20 Seconds:
The Head Teacher starts cleaning up the table and straightening chairs. The children move into the bathroom. On the way to the bathroom, one child asks what they're having for lunch. As she straightens chairs the teacher says, "spaghetti", and goes on with her work.

40 Seconds:
The children are all in the bathroom with the Assistant Teacher. The Head Teacher is setting the table.

60 Seconds:
She continues setting the table. The children start to come out of the bathroom. She says nothing to them.

20 Seconds:
The children are all sitting down chattering and waiting for lunch. The teacher asks everybody to sit quietly for a moment until the food comes. She then brings the plates of food in, telling the children to wait until they all have their plates before starting to eat.

40 Seconds:
The children begin eating. The Head Teacher sits down and begins to eat with them. They eat silently for ten seconds, then the teacher reminds them that they must ask for food instead of grabbing for it.

60 Seconds:
They continue eating. The teacher asks the class whether they know what they're eating. They all shout back "spaghetti."

APPENDIX G

CODING TEST FOR OSCI OBSERVERS

Prepared by

UCLA Head Start Evaluation & Research Center

March, 1969

Part I: List all Context and Content Codes. After you have finished, turn the page, and go through the rest of the test without referring back to page 1.

E & R Center____

Observer_________________________________Date____________

Part II

Class Observation

Note: Fill in record boxes as needed. Be sure that all the children are recorded.

INDOOR	C	OUTDOOR

The teachers are drinking coffee in the kitchen.

Two little boys are fighting on the rug.

Two children are listening to an Xavier Cugat record.

Three children are running around the room making airplane noises and gestures.

One little boy is watching the "airplanes."

Four little girls are having a silent tea party with a volunteer.

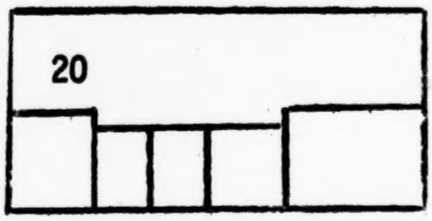
Two children are playing with interlocking blocks.

One little boy is in the bathroom.

TOTAL CHILDREN
PRESENT ______

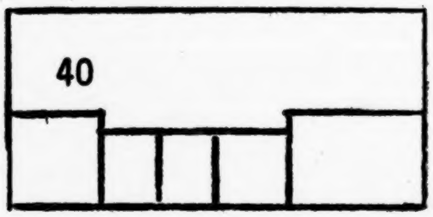
Part III

Head Teacher Observation



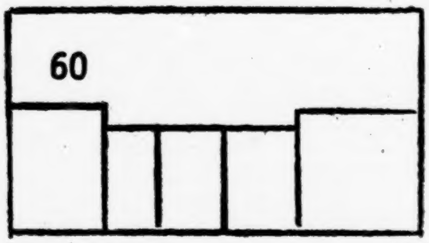
20 seconds:

The Head Teacher and the whole class (15 children) are sitting on the rug. The teacher is pinning shapes on a flannel board and is naming them with the children.



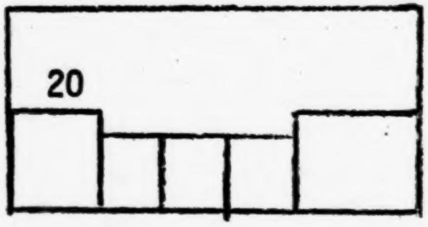
40 seconds:

They continue recognizing shapes. Towards the end of the twenty seconds the teacher says, "Now let's all put our heads down and close our eyes and in a minute we'll go outside."



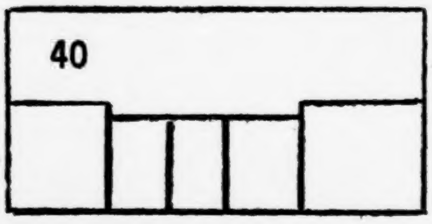
60 seconds:

The children have their heads down. The teacher is moving among them, silently hugging them as they rest.



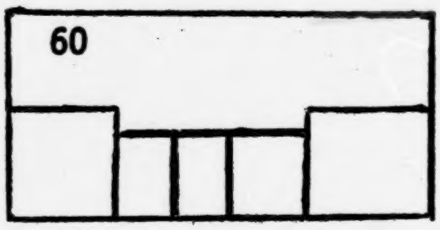
20 seconds:

The children continue resting and the teacher begins to sing softly about "Mr. Sun who is waiting outside."



40 seconds:

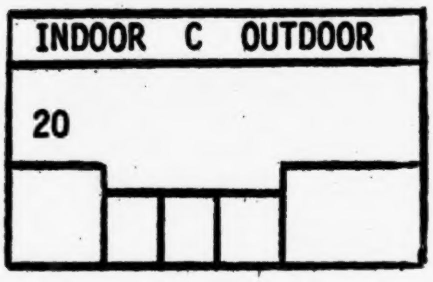
At a nod from the tester, the children begin to get up and line up quietly by the door. The teacher then gets up with them and goes to the door saying, "Let's all get ready and then we'll go outside."



60 seconds:

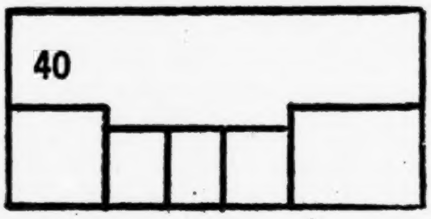
The teacher opens the door and says "Let's all go outside." Everyone runs outside and scatters for the playground toys. The teacher comes out with them and says "What a beautiful day!"

Part III (continued)



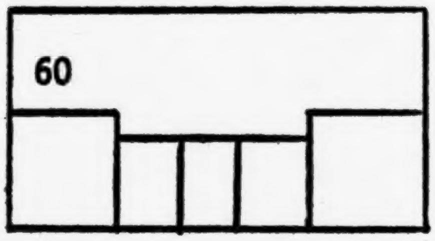
20 seconds:

The children are running, climbing, and riding all over the playground. The teacher moves off to the side and watches them play.



40 seconds:

A ball bounces over to the teacher with two little boys right behind it. As the twenty seconds begin, the teacher picks up the ball, laughing, and begins to throw it back and forth.



60 seconds:

She continues to play ball and laugh. Towards the end of the twenty seconds, she stops and says to the two children, "I have to stop now and go over to see how everyone else is doing."

PART IV

Reports of Research Associated with the Evaluation Study

A. The Development of an Instrument to Measure Teacher Attitudes Toward Evaluation (TATE)

B. Teachers Expectations for Achievement of Children in Head Start (TEACH)

C. Teacher Ratings of Behavioral Objectives as Related to Performance of Children on Specific Tasks

THE DEVELOPMENT OF AN INSTRUMENT TO MEASURE TEACHERS ATTITUDES TOWARD EVALUATION (TATE)

Carolyn Stern and Barbara Rosenquist

Problem

The attitude of a Head Start teacher with reference to the evaluation and research activities carried out in her classroom is an important factor which can affect the validity of the data collected. How the teacher feels about the kinds of instruments used, the nature of the assessments being made, and the presence of observers underfoot, can have an indirect but appreciable impact on the performance of the children as well as on the ease with which the evaluation can be conducted.

In many cases, opposition is due to misunderstanding about the basic purposes of a testing program, and to misconceptions as to how the test scores are to be used. Often teachers may question the value of removing the child from a meaningful learning activity so that he may be subjected to a lengthy testing session. This is especially irritating when the teacher is kept in ignorance of the child's test performance and is thus unable to design curricula to meet the specific needs revealed by the test results.

During the 1967-1968 evaluation program, the UCLA Head Start Evaluation and Research Center had been charged with the responsibility for developing an instrument to describe curricular variations in on-going Head Start classes. In the course of the preliminary explorations, there were many opportunities to talk with teachers with greater freedom than was permissible in the classes which were part of the evaluation sample. It was demonstrated with unexpected consistency that teachers who were suspicious and insecure when they were uninformed about what was occurring in their classrooms were able to relax and make many helpful suggestions after the objectives of the observation were clarified. Almost invariably, these teachers became interested in participating in other research and evaluation studies.

Because the relationship with the operating personnel of the Head Start class can be so critical, the UCLA Head Start Evaluation and Research Center proposed an experimental investigation of an intervention procedure designed to foster favorable attitudes toward, and increased understanding of, the national evaluation program. However, in attempting to determine what criterion measures could be used to test the hypotheses of the study, it soon became apparent that no suitable instruments were available.

While there is a voluminous literature on the subject of teacher attitudes, very few investigators have addressed themselves specifically to teachers' attitudes toward evaluation or research. Torbet (1957) measured the attitude of a group of Colorado secondary school teachers toward informal teacher-made tests, using a projective interview. He reported a definite gap between theory of testing, as taught in the standard course in measurement, and actual practice. The general attitude of the teachers involved was that "testing was an onerous task and an authoritarian weapon." Howard & Berkowitz (1958) measured the reactions of persons to the evaluation of their performances by means of a checker-playing task which supposedly assessed the subject's "effective intelligence." Subjects were given scores reflecting various levels of "favorableness." The results showed that teachers were far more concerned with obtaining reliable and constructive information about their performance than in achieving a "very favorable" rating, per se.

Getzels & Jackson (1963) favor the Minnesota Teacher Attitude Inventory as an instrument which measures with high reliability the attitudes of the teacher toward her pupils as well as the teacher's satisfaction with teaching as a vocation, but the items on this instrument do not include questions about testing. It was therefore necessary to design an instrument which would focus specifically on those attitudes of teachers which relate to the conduct of a comprehensive testing program.

Theoretical Rationale

Most investigators agree that attitudes can be defined as predispositions to respond to a given stimulus or class of stimuli. Further, there is general agreement that attitudes can be characterized as inducing behavior that is either favorable or unfavorable toward the stimulus in question. A way of measuring this type of behavior, called the semantic differential, has been developed by Osgood, Suci, and Tannenbaum (1957). The semantic differential is based on the assumption that the connotative or affective components of meaning can be measured by the rating of objects or ideas with respect to bipolar adjectives.

In the factor analyses of data derived from this semantic differential procedure, three "super" dimensions, Evaluation, Potency, and Activity, are usually found to carry most of the meaning embedded in the concepts. By far the most important of these factors is Evaluation. That is, knowing whether a concept is perceived as "good" or "bad" appears to be the most significant thing one can know about it.

The importance of the Evaluative dimension in measuring meaning has stimulated a review of the work of earlier investigators. Although in agreement as to the definition of attitude, these investigators have implicitly or explicitly objected to a unidimensional view, that is, a single score to represent the totality of the individual's attitude on a particular subject. Allport (1935) felt that although two individuals might be equally favorable toward an object, they might differ in the amounts of favorableness they felt toward various characteristics or components of that object. In the same vein, Chein (1948) agreed that although two individuals might be equally favorable in their attitude toward an object, they might hold different beliefs about what should be done with regard to the object. The multi-component approach has tended to include in a composite description of the attitude toward a particular subject all the various statements that are manifestations or indications of the attitude. While this procedure produced many important insights into the structure of attitudes, it has unfortunately resulted in further confounding the relationship between belief and attitude.

Fishbein and Raven (1962) attempted to clarify these two concepts by providing an operational distinction between "belief" and "attitude." Consistent with the theoretical formulations of Osgood, Suci, and Tannenbaum (1957), attitudes are defined as the evaluative dimension of a concept, e.g., is the concept "good" or "bad?" Similarly, beliefs are defined as the probability dimension of the concept, i.e., is the concept "probable" or "improbable?" Another way of distinguishing these concepts is to characterize attitude as the affective and belief as the cognitive component. From this point of view, the specific definitions of attitude and belief can be held independent of each other. Attitude is not defined as including belief, and belief, defined as a probability dimension, can change independently of attitude. Furthermore, two individuals may differ in belief but have similar attitudes.

Proceeding along these lines, Fishbein & Raven have developed the Attitude and Belief (AB) Scales. In this work they have been led to question whether it is enough to measure the perceived probability of existence of a belief, or whether the precise nature of that existence should also be a concern. This question suggests a distinction between belief in a concept and belief about a concept, the latter being defined as belief in the existence of a number of relationships between the concept and other concepts. It was shown that a change in attitude toward a particular concept could result from a change in belief about that concept. By using the AB scales, the belief in the existence of a stated relationship could be measured. The various beliefs in the relationships between an object and other objects or qualities would then be defined as beliefs about that object. It is obvious that individuals could thus agree in their belief in an object, but differ in their beliefs about that object, i.e., in their estimation of the various qualities and objects which might be associated with a given object.

Following these considerations, Fishbein (1961) recently investigated the functional relationships among attitude toward an object, attitudes toward other concepts, and beliefs about the object, i.e., the beliefs in the existence of relationship between the object and these concepts. Fishbein's theory may essentially be stated as follows: (1) an individual holds many beliefs about any given object, i.e., many different characteristics, attributes, values, goals, and objects are positively or negatively associated with a given object; (2) associated with each of these "related objects" is a mediating evaluative response, i.e., an attitude; (3) these evaluative responses are summative; (4) through the mediation process, the summated evaluative response is associated with the attitude object; and thus (5) on future occasions the attitude object will elicit this summated evaluative response, i.e., this attitude.

According to this theory, an individual's attitude toward any object is a function of his beliefs about the object (i.e., the probability that the object is associated with other objects, concepts, values, or goals) and the evaluative aspect of those beliefs (i.e., the attitude toward the "related objects"). Algebraically, it may be predicted that an individual's attitude toward any object = $\sum_{i=1}^{N} B_i a_i$ where B = belief "i" about the object; a_i = the evaluative aspect of B_i; and N = the number of beliefs. Operationally, there should be a high positive correlation between the predicted attitude toward the object and some obtained measure of that attitude (A_0).

In order to test the above hypothesis, it is necessary to obtain (1) an individual's belief about the attitude object; (2) a measure of each belief and its evaluative aspect; and (3) a measure of attitude toward the object. Fishbein's (1961) data were highly supportive of the hypothesis. This finding, along with the earlier work of Zajonc (1954), Rosenberg (1956, 1960), and others, provided strong support for the general hypothesis that an individual's attitude toward any object is a function of his beliefs about the object and the evaluative aspect of those beliefs. In addition, the study attempted to demonstrate that descriptive or reportorial beliefs about an object are important determinants of an individual's attitude toward that object.

Since a major concern of the UCLA intervention study was the measurement of changes in teacher attitudes and beliefs about evaluation, it was felt that the Fishbein AB techniques would be most appropriate.

Instrument Development

A list of 56 items covering a variety of research and evaluation concerns was constructed and administered to 39 Head Start teachers. The data obtained were analyzed for content coverage, wording, and redundancy. This analysis identified 22 items which could be used as defining the evaluative characteristics, or the A Scale. For each of the A Scale items, which described characteristics of Evaluation, there were four sets of bipolar adjectives: good-bad; unnecessary-necessary; wise-foolish; and dislike-like. The teachers were asked to rate the statements descriptive of evaluation first, so that their responses to the belief statements would not influence their attitudes toward the Evaluation concept.

By adding the words "Evaluation includes" each of the descriptive phrases was made into a positive statement about specific aspects of Evaluation, which constituted the B Scale. The B Scale items were rated in terms of four statements reflecting their probability of occurrence: rarely-frequently; probable-improbable; present-absent; and false-true.

All 44 items were presented on separate sheets in a 3" x 8 1/2" booklet. The instructions for responding to the A Scale were given on the first page of the booklet and for the B Scale on the page immediately following the last item of the A Scale. The final page of the booklet contained the single word "Evaluation" and the four bipolar adjectives of the A Scale, described above. Each teacher was asked to rate Evaluation on the four evaluative dimensions. This last item was included so that there could be a basis for determining whether the descriptive statements developed for this instrument were valid components of an attitude toward Evaluation.

The TATE was pilot tested with 39 subjects, among whom were included Head Start teachers, Child Development Supervisors, and other specialists in early childhood education. As predicted according to the theoretical framework developed by Fishbein, there was a high positive correlation ($r = .54$, $p < .01$) between attitude toward the object ($\sum_{i=1}^{22} B_i a_i$) and the obtained measure of the attitude (A_O, the Evaluation concept). This strong correlation indicates that the statements used in the scale are indeed descriptive characteristics of Evaluation.

RESULTS

Pretest Analysis

A total of 42[1] teachers from the 24 classes assigned to the study were given the TATE at the Orientation Meeting, before they had been informed as to the nature of the intervention in which they would be asked to participate. The analysis of the responses on this pretest indicated that there were distinct patterns of favorable and unfavorable attitudes toward various aspects of the evaluation concept (see Table 1). It was quite clear that teachers had the most positive attitudes toward the activities which provided them with data about a) their children (item #7, R. O. 1; item #16, R. O. 5) with special emphasis on the social-emotional area (item #11, R. O. 3); b) the relationship between their class and the rest of the Head Start program (item #20, R. O. 2; item #13, R. O. 6); and c) the involvement of parents in the education of their children (item #19, R. O. 4; item #18, R. O. 7).

It is interesting to note that the lowest ratings were given when the testing was perceived as non-responsive to teacher needs (item #2, R. O. 22; item #4, R. O. 19) or involved removing children from the classroom (item #10, R. O. 16). The personal threat felt by teachers in an evaluation sample class is evident in the ratings of items having to do directly with the teacher (item #3, R. O. 20; item #6, R. O. 18) or with classroom observation (item #8, R. O. 21; item #22, R. O. 17; item #15, R. O. 15). In the case of item #8, which had the second lowest evaluative rating, observation was seen as having a bad effect on children, whereas in actuality the children for the most part were completely undisturbed by the observers.

In comparing the relative importance of cognitive (R. O. 12) versus affective (R. O. 3) learning experiences, it is clear that most teachers value social-emotional growth above academic achievement. In other words, while realizing that Head Start has the important function of teaching children the cognitive skills required in the public school, the teachers feel that only after the child's feelings of security have been firmly established can any type of academic instruction be effective.

---

[1] While the 24 Head Teachers were present at this meeting, there were only 18 Assistant Teachers. Two classes had not yet been provided with assistants and the other four did not come to the meeting for various reasons.

Except for items #2, 4, 8, & 22, the belief rating was below that of the attitude rating for the same item (see Figure 1). The highest score on the attitude scale (item #7) was 10.9, whereas the highest score on the belief scale (item #20) was 6.21. Perhaps a better comparison is to look at two items which had the same rank order on both scales: item #11, R. O. 3, where the attitude score was 8.54 and the belief score 5.56; and item #16, R. O. 5, where the attitude score was 7.95 and the belief score 5.41. There seems to be a greater distance between attitude and belief scores for the items which obtain the highest ratings; this diminishes in a regular fashion, from a difference of 4.7 points for the top score to only .5 for the second lowest score. An interesting reversal of this trend can be observed on the lowest item, where the attitude rating is a negative score, -1.90, whereas the lowest belief score is .67. These findings tend to support the hypothesis that teachers are not well-informed as to the content or goals of evaluation.

At approximately the mid-point of the year, the group which was receiving the feedback intervention seemed to have reached a high level of rapport with the research staff. It was felt that this enthusiasm might be an ephemeral phenomenon, and that it might diminish considerably with habituation, so that by the time the posttest measure was taken the effectiveness of the intervention would not be measurable. For this reason, it was decided to administer the TATE to the experimental group. Sine the agenda for the feedback meeting was already full, the TATE booklets were broken up and reassembled so as to contain half of the attitude and half of the belief scale items. These were then compiled as alternate forms and randomly presented to the teachers in the experimental group.

Although the data obtained were too limited to warrant statistical analysis, this mid-term evaluation demonstrated that the items rated most positively in the pretest were not affected by the attitude changes which occurred at this time, and those items most closely related to the site variable reflected the greatest amount of change (items #1, 6, 8, 9, and 12). In general, there was an increase in unfavorable attitudes toward observers in classrooms, rating teachers, and effect on child behavior of observers and testers. Positive attitudes toward evaluation were found for increased use of films and classification of classroom activities. With respect to the belief ratings reflecting knowledge of the substantive characteristics of an evaluation program, there was increased awareness that testing of samples of children provides information about all children, and that evaluation was concerned with measuring changes in children.

These mid-year results indicate the beginning of a trend showing that the experimental teachers tend to become more tolerant of evaluation processes which involve interrupting the normal classroom activities, after they have been informed as to the rationale of the measures and have been invited to express their feelings and criticisms about them. However, there is a tendency to become increasingly negative toward the items initially rated most negatively. In addition, the belief ratings begin to show more correspondence with the attitude ratings, increasing for the positive items and decreasing for the items originally rated most negatively.

Posttest Analyses

A. Pre-post changes for total group. Inspection of the relative rank order positions of the items based on the pre and posttest scores reveal that the teachers continue to rate the items having to do with the overall Head Start programs more positively than items having to do with site activities, and items reflecting lack of sensitivity to humanistic considerations most negatively.

The belief ratings increased significantly between the pre and posttest measures for items #13, 19, 20, and 21, and decreased significantly for item #18. While the differences in the remaining items were not significant, nevertheless they did show a trend in the hypothesized direction. Thus these results indicate that a correspondence did develop between attitudes about various activities and the belief that these activities are part of the evaluation program, even though many of the characteristics were specific to the activities of the UCLA Head Start Evaluation and Research Center alone.

B. Treatment differences. In general, the attitude changes for the teachers receiving feedback paralleled those for the total group, significantly increasing for items #7, 18, 19, 20, 21, and 22, and decreasing for items #2 and 4. Additionally, they became significantly more positive about the time spent by observers in the classroom, and did not become more negative toward rating the teacher's performance and the observer's effect on child behavior. These particular differences support the hypothesis that teachers given feedback become less threatened and more positive toward the Evaluation personnel and activities.

On the belief scale, the experimental group significantly changed their beliefs about evaluation on all the items on which the entire group changed. Additionally, their beliefs significantly changed for items 5 (-), 10 (+), 13 (+), and 16 (+). These teachers alone significantly decreased their belief that Evaluation includes spending class time participating in Head Start research, and significantly increased their belief that Evaluation includes removing children from the classroom for testing, knowing how their class compares with other Head Start classes, and testing children. Thus while the total group of teachers regardless of treatment group did show some changes in their beliefs about Evaluation which in fact are congruent with the program, the experimental group changed their beliefs the most and in a direction consistent with the activities of the Evaluation program. Thus these results definitely tend to support the hypothesis that providing feedback to teachers reduces threat, increases their reception to the program, and helps them to distinguish fact from fiction insofar as specific program activities are concerned.

In general the changes for groups C_1 and C_2 paralleled those for the entire group in both the attitude and belief measures, with the major difference being that both these groups showed far less change and on fewer items than the experimental group.

CONCLUSION AND DISCUSSION

From this first use of the TATE in a field study, it has proved to be a useful instrument for revealing areas where attitudes and beliefs about evaluation need to be reconsidered, so that the gap between knowledge and preconception can be diminished. In almost all cases, increases in knowledge were paralleled by increased favorable attitudes toward various components of evaluation. These changes were significantly greater for a group which received feedback during evaluation, but this was true only on the items which were related to the intervention itself. Further use of this instrument in evaluation studies seems to be warranted.

REFERENCES

Allport, G. A Handbook of Social Psychology, Clark University Press, Worchester, Massachusetts, 1935.

Chein, I. Behavior theory and the behavior of attitudes: some critical comments, Psychology Review, 1948, 55, 175-188.

Fishbein, M. An investigation of the relationships between beliefs about an object and the attitude toward that object, Technical Report No. 6, Contract Nonr. 233(54), University of California, Los Angeles, 1961.

Fishbein, M. & Raven, B. The AB scales: an operational definition of belief and attitude, Human Relations, 1962, 15, 35-44.

Getzels, J., & Jackson, P. The teacher's personality and characteristics. In N. L. Gage (Ed.), Handbook of Research on Teaching. Rand McNally & Co., Chicago, 1963.

Howard, C., & Berkowitz, L. Reactions to the evaluations of one's performance, Journal of Personality, 1958, 26, 494-507.

Husek, T. The measurement of meaning for travels in semantic space, Educator, University of California, Los Angeles, October, 1963.

Osgood, C., Suci, G. & Tannenbaum, P. The Measurement of Meaning, University of Illinois Press, Urbana, 1957.

Rosenberg, M. Cognitive structure and attitudinal affect, Journal of Abnormal Social Psychology, 1956, 53, 367-372.

Rosenberg, M. An analysis of affective-cognitive consistency, in Attitude Organization and Change, Yale University Press, New Haven, 1960.

Torbet, D. The attitude of a select group of Colorado secondary school teachers toward informal teacher made tests as measured by a projective interview, Journal of Educational Research, 1957, 50, 691-700.

Zajonc, R. Structure of the cognitive field, Unpublished Doctoral Dissertation, University of Michigan, 1954.

Table 1

Mean Pretest and Posttest Scores and Rank Order of Items on Attitude (A) and Belief (B) Scales of TATE

Item	Pretest (N=42)				Posttest (N=40)			
	Mean Score		Rank Order		Mean Score		Rank Order	
	A	B	A	B	A	B	A	B
1. The time spent by observers in the classroom.	6.74	5.49	8	4	7.73	2.73	5	17
2. Waiting a long time before getting test results	-1.90	4.08	22	13	-2.10	4.30	22	9.5
3. Relating job security to teacher ratings.	3.41	0.67	20	22	3.18	1.85	20	19
4. Testing children in a few classes to find out about Head Start children in general.	4.10	4.56	19	10	0.58	3.03	21	15
5. Spending my class time participating in Head Start research	6.31	2.82	11	18	4.03	1.28	19	20
6. Rating the teacher's performance	4.43	3.41	18	15.5	5.88	4.30	15	9.5
7. Providing teachers with information about test results.	10.9	4.46	1	11	9.45	4.25	2	11
8. The observer's effect on child behavior.	2.72	4.38	21	12	4.23	0.67	18	22
9. Filming classroom activity.	6.23	3.64	12	14	7.08	0.73	10	21
10. Removing children from the classroom for testing.	5.13	4.64	16	8.5	7.05	6.00	11	2
11. Emphasizing the children's social-emotional growth.	8.54	5.56	3	3	8.23	4.60	4	7

Table 1 (cont'd.)

Item	Pretest (N=42)				Posttest (N=40)			
	Mean Score		Rank Order		Mean Score		Rank Order	
	A	B	A	B	A	B	A	B
12. Classifying classroom activities.	6.31	4.64	10	8.5	6.55	3.55	13	13
13. Knowing how my class compares with other Head Start classes.	7.69	2.74	6	19	7.30	5.38	9	4.5
14. Emphasizing the children's academic achievement.	6.10	3.41	13	15.5	5.80	3.20	16	14
15. Observing the teacher at work.	5.26	5.10	5	6	7.57	6.03	7	1
16. Testing children.	7.95	5.41	5	5	7.70	5.38	6	4.5
17. Spending my out-of-class time participating in Head Start research.	5.56	2.33	14	20	5.58	2.80	17	16
18. Spending my time helping parents teach their children at home.	7.28	2.21	7	21	6.28	2.48	14	18
19. Spending my time giving parents information.	8.08	2.87	4	17	8.83	3.63	3	12
20. Knowing the relationship between the child development center and the national Head Start program.	8.74	6.21	2	1	9.75	4.38	1	8
21. Specifying goals for changing child behavior.	6.67	4.69	9	7	7.38	4.73	8	6
22. Having observers in the classroom.	4.46	5.79	17	2	6.65	5.50	12	3

Table 2

T-test Values for Differences between Mean Scores on Pretest and Posttest for Attitude and Belief Scales (by Treatment Group)

Item	Attitude				Belief			
	E N=14	C-1 N=9	C-2 N=7	Total N=30	E N=14	C-1 N=9	C-2 N=7	Total N=30
1	5.8*	0.0	0.7	2.9	1.2	2.3	-4.9	0.1
2	-15.3**	-12.2**	-9.9*	-13.1**	-1.6	-0.1	5.0	0.4
3	1.2	2.7	2.3	1.9	-1.1	-1.2	-0.7	-1.0
4	-5.1*	-8.1*	-6.1**	-6.2**	-1.5	-1.4	3.0	-4.4
5	1.1	-4.4	-1.4	-1.1	-4.3*	-1.9	0.3	-2.5
6	-4.1	-0.6	-4.3*	-3.1*	-2.6	0.1	0.0	-1.2
7	7.1**	1.6*	-1.3	3.5*	1.4	-1.8	1.1	0.4
8	-1.9	-3.0	-6.6*	-3.3*	-2.6	-0.7	-1.0	-1.6
9	1.1	2.4	-2.4	0.7	-2.1	-1.1	-0.4	-1.4
10	3.9	2.0	0.3	2.5	6.2**	-2.9	0.1	2.1
11	0.9	0.0	-2.4	-0.1	2.8	-1.9	-2.3	0.2
12	2.0	0.0	3.0	1.6	-0.1	2.2	3.7	1.5
13	0.4	1.2	1.9	1.0	6.1**	2.8	4.7	4.8**
14	-3.1	-2.3	1.7	-1.8	-2.0	1.4	7.0	1.1
15	-0.1	-0.6	-2.1	-0.7	3.6	0.7	-3.6	1.1
16	-0.1	1.7	3.9	1.4	4.8*	0.1	-2.3	1.7
17	2.6	1.7	-1.6	1.4	-2.1	-4.1*	-1.3	-2.5
18	20.9**	23.8**	23.3**	22.3**	-6.3*	-4.1*	26.3**	-5.7**
19	24.5**	23.8**	26.6**	24.8**	19.2**	20.1**	22.4**	20.2**
20	25.8**	25.7**	27.4**	26.1**	20.4**	20.6**	22.3**	20.9**
21	24.0**	25.2**	22.9**	24.1**	20.6**	20.9**	23.6**	21.4**
22	22.3**	23.4**	21.0**	22.3**	22.4**	20.4**	22.7**	21.9**

$^{*}p < .05$ $^{**}p < .01$

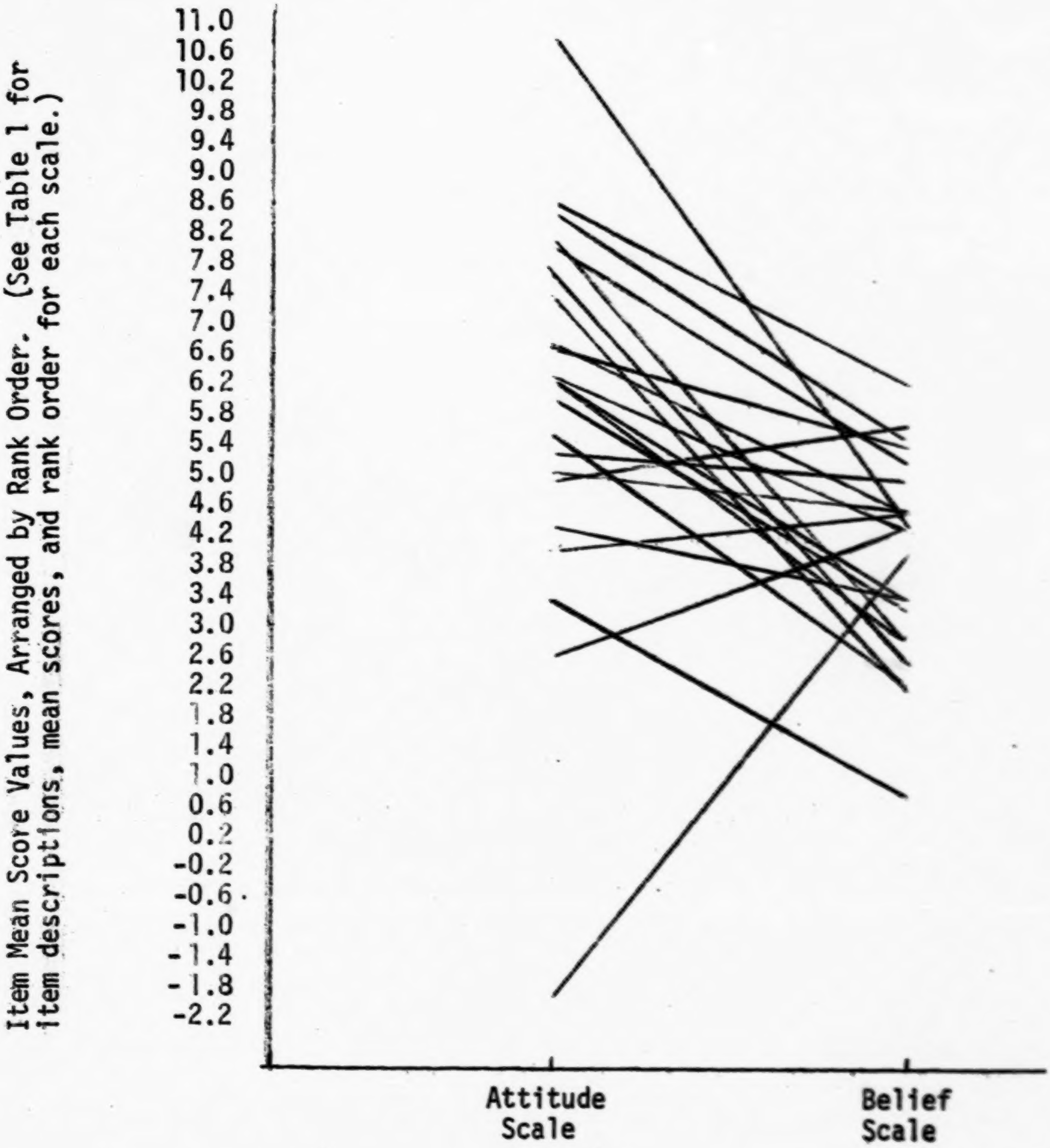


Figure 1. Comparison of Rank Order of Pretest Scores on Attitude and Belief Scales of TATE.

TEACHERS EXPECTATIONS FOR ACHIEVEMENT OF CHILDREN IN HEAD START (TEACH)

Carolyn Stern, Harriet Prichard, and Barbara Rosenquist

Introduction

During the past few years the concept of accountability in federal funding of educational projects has been paired with the emphasis on behaviorally-based criteria for measuring success in teaching. This promising marriage has led to the propagation of numerous pre-post evaluations of innovative school programs. Such designs are viable in the elementary grades, where there is some degree of correspondence both between the operational definition of the program goals and the instruments used to measure level of achievement, and between the goals expressed by the teachers and the types of behaviors measured by the evaluation instruments. Unfortunately, there is little evidence that either of these necessary correspondences exists in the early years of schooling.

In the first place, there is a serious dichotomy between the emphasis on affective-emotional-social behaviors which nursery school teachers and child development specialists conceive as appropriate for the four- to six-year-old child, and the current popular focus on remediating cognitive deficits and instilling pre-academic skills. There is also a wide discrepancy between the professed goals of the classroom practitioner and the activities which are instituted presumably to achieve these goals. Finally, there is a notorious lack of instruments for assessing the achievement of goals, even when they are expressed in measurable terms. It is thus not too surprising that several highly publicized evaluations of preschool programs reveal either ephemeral or non-significant gains in concrete academic skills while the teacher places her values on achievements in the affective domain, which remains comfortably uncharted. This is somewhat analogous to using gain in height to establish the effectiveness of an experimental nutrition program: While there is undoubtedly a physiological connection between food and growth, a valid test requires far more direct and specific criteria.

Instrument Development

The need for an instrument which would be able to relate teacher goals, classroom activities, and achievement in children became apparent within the context of the UCLA Head Start Evaluation Center's Feedback Study. It was the hypothesis of this investigation that there would be a relationship between these three variables such that providing evaluation information to teachers would produce changes in classroom procedures as well as in value systems. To test this hypothesis, it was necessary to develop a comprehensive instrument which could represent the wide variety of goals valued by educators from extremely divergent pedagogical persuasions.

The first step in this instrument development was to amass a pool of value statements. To this end, an intensive search of the literature was carried out and hundreds of items culled from check lists, attitude inventories, and other teacher measures. In addition, teachers, supervisors, and early childhood specialists were asked to provide descriptions of valid goals for the young child. All these were typed on individual cards and reviewed at a meeting of teachers and specialists on the staff of the UCLA Head Start Evaluation Center. The items were tentatively placed in traditional categories such as motor skills, social and emotional development, self-care, etc., and redundant and overlapping statements were eliminated.

The next step was to devise descriptions of behaviors which would be accepted as evidence of the achievement of the stated goal. In this process many goal statements expressed as "consideration for others," "positive self-concept," and "intellectual curiosity" were translated into several items representing specific behaviors for each of these categories, for example, "waits his turn in the use of playground equipment," for the first, "smiles when looking at himself in the mirror" and "identifies positively with his own race," for the second, and "asks questions about objects in the environment" and "explores different solutions to a problem" for the third. After informal discussion with colleagues and others in the field of early education, the original set of behavioral statements was winnowed to a list of 100 items.

The third step in the instrument development was devising the format for the item presentation. It was felt that a straightforward rating scale type of inventory would not get at the question of correspondence between the value placed on a particular behavior and the expectation of achieving that behavior within the constraints of the preschool classroom. Fortunately, at this point Dr. Edward Suchman, the consultant for the major evaluation study, was visiting the UCLA Head Start Center. He suggested using a model he had applied in an unpublished study of teacher attitudes toward the poverty group and their exposure to such groups. This approach served as the inspiration for the development of a format which requires two separate responses to the same set of items: in the first time through the list, the respondent is asked to indicate the degree of importance attributed to a particular item; the second time through the rating is for expected level of performance.

Table 1 includes the instructions given to respondents and a sample of the item format. This form of the instrument was filled out by a heterogeneous group of respondents, including the national Evaluation Coordinator, and the data subjected to a principle factor solution, using BMD Program X72, at the UCLA Campus Computing Network. Orthogonal rotation using the varimax method developed the following seven factors: 1) Curiosity (15 items); 2) Independence (15 items); 3) Social Awareness (8 items); 4) Sensory and Cognitive Skills (29 items); 5) Motor Skills (6 items); 6) Self-Awareness (5 items); and 7) School-Appropriate Behavior (22 items).

It was expected that as a result of this pilot testing of the TEACH and the subsequent factor analysis it would be possible to substantially reduce the number of items. However, the need to use the instrument in the pretesting phase of the evaluation study and the delay in obtaining the results of the factor analysis left insufficient time to prepare a modified form. Thus the original instrument, although admittedly overlong, was used in the UCLA Feedback Intervention.

Application of the Instrument

The UCLA Feedback Intervention involved 24 Head Start classes in a primarily urban setting. Both the Head Teacher and the Assistant Teacher were asked to fill out the TEACH as a pretest measure during the first orientation meeting. At this time they were asked to rate the items in terms of (a) degree of importance or relevance for Head Start children; and (b) level of expectation of achievement for that particular objective. The same measure was administered at the final meeting which took place approximately eight months later.

It was hypothesized that, compared to the control group, teachers given feedback would show a higher degree of correspondence at the end of the year between their ratings of the importance of certain behavioral goals for Head Start children and their estimates of achievement of these goals. Also, there would be a tendency to place greater value on those goals they felt they had achieved compared to goals with which they had had little success.

Results

Table 2 presents the pretest mean scores for the total group, with each mean having a range from 0 (low) to 9 (high). Using the seven factor structure obtained in the preliminary analysis, the item means were summed to derive a grand mean for each factor. These means for the total sample and for each of the treatment groups are presented in Table 3.

For the total group of teachers, the means on the Importance Scale were highest for Independence, Curiosity, Social Awareness, and Self-Awareness, in that order. They were decreasingly lower for Motor Skills, Sensory and Cognitive Skills, and School Appropriate Behavior. These results indicate considerable face validity in terms of the fact that Head Start teachers apparently view the development of emotional, social, and situational maturity as more important than academic skills *per se*.

Interestingly enough, the Expectation scores are generally higher than the Importance ratings. Several speculations may be advanced for these findings. The teachers may stress the behaviors for which they hold the highest importance, thus a type of self-fulfilling prophecy could be operative. If this is true, there should be lower Expectation ratings for those factors which are rated low on the Importance Scale. While the rank order of the Expectation ratings follows the same direction as the Importance ratings, the numerically higher Expectation ratings seem to contradict this assumption. Another possibility is that while the teachers do in fact view readiness for learning as a prime function of Head Start, they feel that this goal can best be achieved by first establishing strengths in such non-cognitive areas as basic trust and self-esteem.

Overall, the rank ordering on the factors for the three treatment groups showed the same trend as for the total group. Developmental or readiness factors have higher rank orders than educational factors, and Expectation ratings are higher than Importance ratings.

After the posttest data were obtained, a series of t-tests was run between the pre- and posttest scores on each item. The results of this analysis are presented in Table 2. There were no statistically-significant changes between the pretest-posttest Importance ratings. In other words, the teachers included in this sample had very stable feelings about their objectives. The speculation that there might be a decrement in the value placed upon an objective when it was not adequately achieved, was not supported. The differences that did obtain were between the expectations for achievement and the ratings of level of achievement obtained. Without exception, these differences were in terms of a decrease in mean scores. Thus it appears that for the total group of teachers, the degree of correspondence between Importance and Expectation increased during the year:

While greater variance was found among the teachers in the experimental treatment, the number of items was so large and the sample size of teachers so small that it would have been meaningless to look for statistical differences between the pretest-posttest scores by treatment group. An informal comparison was done on the teachers included in the TEACH-TASK analysis (Stern & Rosenquist, 1970). Inspection of the individual scores indicated practically no difference for any of the scale comparisons (importance-expectation, pre; importance-achievement, post) and the differences that did exist did not indicate any consistent trend.

Discussion and Conclusion

While the overall results tend to support the hypothesis of correspondence, nevertheless several problems appear inherent in the TEACH instrument. By far the most important is its excessive length, which many teachers found onerous. This fatigue effect expressed itself in a tendency to respond unthinkingly, thus little differentiation either among items or factors was obtained. Also, the large number of rating levels made it difficult to differentiate among the items; almost all the teachers tended to respond in the high direction. Thus it appears that while the general rationale of the TEACH has merit, as supported by the data analysis, additional work is needed in the direction of reducing the number of items and response categories.

The TEACH was also administered to 34 teachers and 15 parents in a research study which was carried out concurrently with the Evaluation (Stern, Kitano, Gaal, Goetz, & Ruble, 1970). However, the data was analyzed separately for that study, using only 95 items with 10 factors.

An adaptation of the TEACH was made for a new research study concerned with parent involvement. The new instrument (PEACH) has 77 items and only four response categories. The responses on the pretest have been used as jumping off points for the group meetings which constitute the experimental treatment. Thus, while the discriminative ability of the instrument has yet to be established, it does offer many insights into differences in value systems which are providing the basis for meaningful group discussions in this new study.

Table 1

TEACHERS EXPECTATIONS FOR ACHIEVEMENT OF CHILDREN IN HEAD START (TEACH)

Even among educators and specialists in child development, there is difference of opinion as to appropriate goals and activities for children in Head Start. We have collected a large number of items describing a variety of behaviors and would like you to express your reactions to each of them.

In Scale I, indicate to what extent you think that a particular behavior or activity is important and should be included in Head Start. For instance, if you think that "sits quietly at the table during mealtimes" is not at all important, you would give it a "low" rating of 0; if you think it extremely important you would give it a "high" rating of 9. If you feel that it falls somewhere between these two extremes, you can give it any rating between 0 and 9. Ask yourself as you mark each item: "How important is this goal for the average child in my class?"

Remember, this is not a test with right or wrong answers. There are equally good teachers who hold quite opposite opinions. Also, all your responses will be kept completely confidential, so do not hesitate to express ideas which you may feel are different from those of other teachers and child development specialists.

After you have rated each item on Scale I in terms of importance, we would like you to rate on Scale II the extent to which you feel your children will acquire the described behavior as a result of their Head Start experience. It is quite possible that a skill is very important, but so difficult to teach that you probably won't get very far in one year of Head Start. On the other hand, you may judge an activity to be relatively unimportant, but easy to teach.

NOTE: Fill in Scale I for all items. Then fold paper so that Scale II on back of page matches the same items on the front of the page in the same position as Scale I.

EXAMPLES

Item	Scale I Degree of Importance										**Scale II** **Expectation of Learning**									
	Low					High					Low					High				
	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
51. Names numerals from 1 to 10																				
52. Sits quietly at the table during mealtime																				
53. Takes responsibility for his own mistakes																				
54. Volunteers ideas of his own																				
55. Recites the entire alphabet by memory																				

Table 2

Mean Pre- and Posttest Scores and Differences between Importance and Expectation[1] Ratings by Factors and Items

Factor		Pretest Mean Scores				Test of Difference between pre- and posttest ratings	
		Importance N = 38		Expectation N = 38		Importance N = 28	Expectation N = 28
	Item No.	Mean	S.D.	Mean	S.D.	t-value	t-value
I: Creativity	15	7.55	2.32	8.10	1.20	0.33	-0.71*
	20	7.53	1.95	7.79	1.53	-0.39	-0.07
	22	6.55	1.90	7.11	2.25	0.11	-0.19
	23	7.16	1.89	7.49	2.05	-0.14	-0.64*
	25	7.63	1.71	7.42	2.53	0.04	-0.93*
	30	6.87	2.24	7.42	2.26	0.11	-0.53
	54	6.89	2.41	6.71	2.84	0.46	-0.11
	56	5.42	2.31	5.68	2.48	-0.11	-0.26
	59	5.71	2.48	6.55	2.37	0.41	-0.11
	70	5.68	2.41	6.60	2.44	-0.39	-0.93
	72	7.95	1.69	7.87	2.00	0.11	-0.39
	79	6.84	2.12	7.03	2.41	-0.26	0.11
	84	5.99	1.95	6.99	1.69	0.0	-0.50
	98	6.58	2.56	6.97	2.52	0.04	-0.31
	100	8.16	1.50	8.34	1.07	-0.36	0.32
II: Independence	2	6.34	2.77	7.39	2.21	-0.25	-1.00**
	8	6.89	1.98	8.05	1.43	-0.11	-0.36
	29	8.21	1.51	8.26	1.13	-0.18	-0.43*
	39	6.99	2.21	6.76	2.77	-0.25	-0.50
	48	6.26	2.62	7.29	2.36	-0.32	-1.22**
	53	6.15	2.34	6.16	2.44	-0.25	-0.75*
	57	6.87	2.16	6.63	2.92	-0.43	-0.74
	62	6.63	2.34	7.31	2.41	-0.43	-0.46
	63	6.95	2.48	6.84	2.88	-0.44	-0.96*
	73	7.18	2.50	7.37	2.42	-0.41	-0.74
	88	5.71	2.73	6.42	2.28	-0.57	-0.64
	89	7.29	2.11	8.03	1.57	0.36	-0.18
	94	7.24	2.06	6.95	2.31	-0.32	-0.04
	96	6.87	2.35	6.84	2.52	0.11	-0.31
	97	7.16	2.02	7.49	1.55	-0.57	-0.82*

[1] On the Posttest, the Expectation Scale was expressed as estimate of achievement

*$p < .05$

**$p < .01$

Table 2 (cont'd.)

Factor		Pretest Mean Scores				Test of Difference between pre- and posttest ratings	
		Importance N = 38		Expectation N = 38		Importance N = 28	Expectation N = 28
	Item No.	Mean	S.D.	Mean	S.D.	t-value	t-value
Social Awareness	9	5.89	2.68	6.74	2.24	0.52	-0.61
	33	7.52	1.70	7.49	1.95	0.18	0.21
	35	6.71	2.47	6.87	2.56	-0.04	-0.14
	38	7.42	1.83	7.29	2.28	0.14	0.14
	44	7.31	2.09	7.49	2.47	0.14	-0.43*
	58	6.89	2.54	7.03	2.59	-0.79	-0.79
	78	6.89	2.04	7.29	2.23	-0.14	-0.14
	86	6.63	2.53	7.68	1.53	-0.03	-0.04
Sensory and Cognitive Skills	1	4.16	2.89	5.65	2.81	0.32	-0.32
	3	5.76	3.07	7.61	1.97	0.65	-0.46
	6	5.39	2.81	6.05	2.72	-0.46	-0.27
	7	6.92	2.32	8.05	1.66	0.30	0.04
	19	5.81	2.94	6.37	2.54	-0.96	-0.57
	27	6.18	2.68	7.03	2.63	0.0	-1.22*
	31	7.50	1.90	7.29	2.39	-0.39	-0.79**
	34	7.66	1.73	7.61	2.54	-0.18	-0.36*
	37	7.39	2.28	7.79	2.08	0.54	-0.29
	41	6.21	2.26	7.10	2.37	-0.11	-0.54
	43	6.26	2.55	7.03	2.75	-0.25	-0.48**
	45	5.03	2.96	6.05	2.47	-0.79	-1.64**
	46	6.74	2.33	6.84	2.34	-0.54	-0.54
	47	6.18	2.49	6.76	2.65	-0.07	-1.14*
	49	4.31	2.62	5.08	2.90	-1.07	-1.54*
	50	5.45	2.59	5.63	2.54	-0.54	-1.14
	51	6.81	2.52	7.21	2.71	-0.46	-0.32
	60	6.58	2.84	7.45	2.42	0.21	-0.54
	67	6.08	2.42	6.81	2.31	0.18	-0.36
	71	5.42	2.80	6.92	2.64	-0.52	-1.32**
	74	7.24	1.81	7.05	2.31	-0.32	-0.32
	76	6.37	2.86	7.10	2.24	0.25	-0.21
	81	6.37	2.58	7.16	2.20	-0.18	-0.93*
	82	5.68	2.63	6.79	2.17	0.39	-0.18
	87	5.71	2.36	6.39	2.63	0.55	-0.26
	90	6.29	2.24	6.92	1.92	-0.57	-1.00
	91	6.34	2.59	7.37	1.58	-0.82	-0.75
	92	7.26	2.42	7.37	2.88	0.0	-0.81*
	99	5.42	2.58	6.24	2.40	-0.36	-1.07*

$^{*}p < .05$

$^{**}p < .01$

Table 2 (cont'd.)

Factor		Pretest Mean Scores				Test of Difference between pre- and posttest ratings	
		Importance N = 38		Expectation N = 38		Importance N = 28	Expectation N = 28
	Item No.	Mean	S.D.	Mean	S.D.	t-value	t-value
Motor Skills	4	5.26	2.87	6.92	2.37	0.24	-0.18
	26	4.99	2.94	6.26	3.00	0.04	-1.88**
	36	6.79	2.85	7.18	3.05	0.28	0.24
	66	5.79	2.59	6.47	2.82	0.50	0.22
	77	6.08	2.58	6.81	2.37	-0.22	-0.36
Self Awareness	12	7.29	2.08	7.37	2.10	0.64	0.54
	13	6.13	2.39	6.31	2.51	0.46	-0.15
	14	6.66	2.58	7.10	1.98	-0.15	-0.11
	61	5.39	2.54	6.05	2.87	0.07	-0.29
	83	6.99	2.28	7.39	2.28	0.07	-0.78**
School-Appropriate Behaviors	5	6.11	2.88	6.79	2.55	0.25	-0.32
	10	4.16	3.11	5.50	2.77	0.21	-0.11
	11	7.24	2.51	8.08	1.71	0.07	-0.71
	16	6.63	2.47	6.81	2.42	0.18	-0.25
	17	3.55	3.21	4.87	3.09	-0.36	-0.46
	18	5.21	2.92	6.24	2.71	-0.61	-0.68
	21	4.45	2.30	6.16	2.27	0.18	0.18
	24	5.50	2.72	6.55	2.83	0.29	-1.19*
	28	5.79	2.51	6.31	2.69	0.43	-0.29
	32	3.21	2.46	4.16	2.98	-1.00	-1.00
	40	6.18	2.32	6.97	1.99	0.21	-0.21
	42	6.16	2.70	7.24	2.37	-0.25	-0.67*
	52	3.21	2.73	4.29	3.05	0.21	-0.04
	55	2.87	2.47	3.99	3.08	0.25	0.11
	64	3.99	2.99	4.89	2.87	-0.86	-1.25**
	65	4.61	2.87	6.08	2.89	0.21	-0.19
	68	5.16	2.66	6.26	2.61	-0.79	-0.79
	69	3.66	2.70	4.97	3.13	-0.18	-0.53
	75	4.31	3.19	5.39	2.95	-0.39	-0.32
	80	4.16	2.80	5.16	3.02	0.50	-0.03
	85	6.26	2.27	7.29	2.46	0.29	-0.07
	93	5.31	2.83	6.66	2.47	-0.07	-0.32

$*p < .05$

$**p < .01$

Table 3

Rank Order of Means for 7 Factors on the
Teacher Expectations of Achievement of Children in Head Start (TEACH) Scale
(by Treatment Groups)

Factor	Scale	Experimental N = 15		Control 1 N = 11		Control 2 N = 12		Total N = 38	
		Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
I. Curiosity (15 items)	Importance	6.87	2.5	6.42	3	7.63	1	6.84	2
	Expectation	7.06	1.5	8.09	1	7.83	1	7.55	1
II. Independence (15 items)	Importance	7.05	1	6.93	1	7.15	3	7.03	1
	Expectation	7.13	3	7.81	4	7.72	2	7.51	2
III. Social Awareness (8 items)	Importance	6.87	2.5	6.66	2	7.52	2	6.94	3
	Expectation	7.06	1.5	7.65	5	7.61	4	7.40	4
IV. Sensory and Cognitive Skills (29 items)	Importance	6.27	6	5.18	6	6.25	4.5	5.90	6
	Expectation	6.94	4	7.84	3	7.49	5	7.41	3
V. Motor Skills (6 items)	Importance	6.40	5	5.60	5	6.25	4.5	6.05	5
	Expectation	6.88	5	7.21	6	7.69	3	7.23	5
VI. Self-Awareness (5 items)	Importance	6.66	4	6.33	4	6.24	6	6.52	4
	Expectation	6.53	6	7.94	2	7.24	6	7.18	6
VII. School-Appropriate Behavior (22 items)	Importance	4.84	7	4.77	7	4.91	7	4.83	7
	Expectation	5.78	7	6.73	7	6.25	7	6.23	7

TEACHER RATINGS OF BEHAVIORAL OBJECTIVES AS RELATED TO PERFORMANCE OF CHILDREN ON SPECIFIC TASKS

Introduction

One of the main criticisms of achievement testing, especially in the field of early childhood education, is that teachers value affective (social-emotional) behaviors whereas most tests measure cognitive skills. Thus a valid basis for the negative findings of the Westinghouse report (1969) might be expressed in terms of the lack of correspondence between the objectives of Head Start teachers and the types of instruments used in the evaluation. For example, many traditional preschool teachers couldn't care less if their children didn't learn to recognize letters of the alphabet, yet this skill is one of the criteria on which the success of Head Start may be based.

In the UCLA Head Start Feedback Intervention Study, it was important to determine the types of values teachers placed on specific behaviors, and to what extent they felt they could help children acquire these skills over the school year. The instrument developed to measure these values was the TEACH (Teachers Expectations of Achievement in Head Start) scale (Stern, Prichard, and Rosenquist, 1970). It was hypothesized that changes in teachers' ratings of a specific behavioral objective would be related to the degree to which they felt the children had succeeded in acquiring the behavior. To obtain a valid measure to compare with teacher ratings on the TEACH items, it was necessary to construct an instrument which would measure child behavior on item-related tasks.

A frequency analysis was run on the pretest importance ratings of the TEACH items for the total intervention sample, using the UCLA BIMED Computer program, to determine which items discriminated among teachers, i.e., which items were rated high in importance by some teachers and low by others. From the high-discrimination items, only those most readily expressed in terms of a behavioral task were used in developing a large pool of potential test items. In designing the tasks, the teachers in the feedback group were asked for suggestions as to what they would accept as evidence of the achievement of a particular objective. Their suggestions and comments proved very helpful and informative. For instance, they pointed up the difficulty of attempting to devise a comprehensive set of objective-related tasks for such items as "Uses correctly newly acquired vocabulary." While a teacher who rated this high in importance might actually have achieved this objective, it would be impossible to confirm as different children might have acquired different words. Since an overriding consideration in designing the instrument was to minimize the time children were out of the classroom, only a sampling of items could be used. In deciding which items should be retained, those which required extensive observation, e.g., "Puts materials away after an activity without being told," or "Asks questions about objects in the environment," were eliminated.

With the above considerations in mind, tasks related to 25 of the 100 TEACH items were selected for a pilot instrument. The test items and the number of the related TEACH items are shown on pages 177 to 178.

METHOD

Subjects

One of the Delegate Agencies in the 1968-69 Feedback study agreed to cooperate in pilot testing the instrument. This agency had three school sites, one in each of the three treatment groups, or a total of seven classes. The agency was about average in terms of teacher turnover. Five children in each of the seven classes were tested.

Procedure

All children were tested individually, using the same test items regardless of which ones the teachers had rated high in importance. This would make it possible to test not only whether or not there was correspondence between teachers' ratings and children's performance, but also whether there was greater correspondence at the high-high (high teacher rating-high pupil achievement) than in the low-low (low teacher rating-low pupil achievement) level.

A mean score on the tasks for each TEACH item was obtained for each class. These scores were then compared to the rating of the teacher of that class for that item. Pre and posttest scores were computed for both the importance scale and the expectation[1] scale to determine the level of correspondence between actual class performance and the rating given by the teacher of that class to a particular goal.

RESULTS AND DISCUSSION

For any particular item, there was very little difference in the ratings of teachers either between the two scales on pre or posttest or for the same scale from pre to posttest. The importance ratings were slightly higher than the expectation ratings, with generally lower expectation ratings on the pretest than achievement estimates on the posttest. It appears as though the teachers had underestimated the capabilities of the students or their own capacity to produce change. Because of the generally low difference between the four scales, to increase reliability the teacher score used for the comparison of teacher ratings with children's performance was the mean of the four ratings.

---

[1] On the posttest, the "expectation" scale was changed to an "estimate of achievement" scale.

A rough standardizing procedure based on a total of 100 was used to equate class mean score with teacher (Head Teacher and/or assistant teacher) mean score for each item. Each score was then categorized as high (70-100), medium (50-70), and low (less than 50). Thus, there could be correspondence between teacher ratings and class performance at the "high-high," "medium-medium," and "low-low" levels.

In general there was a linear relationship between the teachers' mean scores and the children's mean scores, supporting the hypothesis that the children achieved highest on those items which the teachers had rated high in importance and lowest on those rated low in importance. The percentage of correspondence ranged from 96% agreement to a low of 32% agreement for some teachers.

The instances of low correspondence most often occurred because of a general tendency to rate all items in the high direction. In other words, children rarely scored higher than the teachers' ratings but often scored low when the teachers scored high. This finding indicates that it might be more worthwhile to have fewer rating categories, forcing the teachers to differentiate more clearly among the items.

Additional information about the teachers was obtained from the data aides who had monitored the various sites. It was found that those teachers who were rated as "poor" by the monitors tended to show lowest item differentiation, ranking most items as being high in importance. This tended to decrease the correspondence between their ratings and class performance.

Several additional problems became apparent during the pilot test.

1. The large rate of teacher turnover in the local Head Start programs often made it impossible to obtain valid pre-post measures.

2. Teachers generally tended to place high value on affective behaviors, which are extremely difficult to evaluate objectively, especially in a single-session test.

3. An instrument which asks teachers to scale individual items makes it difficult for teachers to express comparative values among items. Perhaps a more effective measure would be to give each teacher a number of objectives to order in a form of Q-sort.

In spite of the problems indicated, in general the results of this pilot study demonstrated that there is a correspondence between what teachers value and what children learn. This provides a substantive basis for the inferences drawn from the analysis of the 1966-1967 Head Start evaluation, where low average gain scores were found on instruments measuring primarily cognitive skills, and teachers' ratings of educational objectives favored social-emotional development.

UCLA HEAD START EVALUATION AND RESEARCH TEACH TASK - JUNE 1969

CHILD'S NAME ______________________ SCHOOL ____________ CLASS ______________

CHILD'S ID ______________ DATE _________________ EXAMINER ________________

Instruction	Response	Related TEACH Item #
1. Point to the car behind the plane.	R W DK	3
2. Point to the car in front of the truck.	R W DK	
3. Point to the first one.	R W DK	
4. Point to the last one.	R W DK	
5a. Listen: Pig, boat, hat. Which one sounds like big?	R W DK	6
b. Listen: Shoe, chair, fly. Which one sounds like high?	R W DK	
c. Listen: Cat, horn, bear. Which one sounds like mat?	R W DK	
d. Listen: Tree, man, bike. Which one sounds like tan?	R W DK	
6a. Which one looks like you? (Like sex, 3 ethnic types)	R W DK	14
b. Which one do you like best?	R W DK	
7. Tell me the story about Jack and Jill.	______________	18
8a. What does this tell us? (Cardboard clock)	R W DK	32
b. What time is it? (Hands set at 1 o'clock)	R W DK	
c. What time is it now? (Hands set at 3 o'clock)	R W DK	
9a. Put your hand in this bag and find something soft. (Cotton and sandpaper)	R W DK	46
b. Hand me the circle. (Wood circle and square)	R W DK	
c. Hand me something small. (Large and small squares)	R W DK	
10a. What number is this? (2)	R W DK	51
b. What number is this? (4)	R W DK	
c. What number is this? (3)	R W DK	
d. What number is this? (1)	R W DK	
e. What number is this? (5)	R W DK	
11. Say the alphabet for me. (4 consecutive letters)	R W DK	55
12. Tell me a story about these pictures. (Kite-bicycle)	________	59
13 What does this say? (Write child's first name)	R W DK	60
14a. What is this taste? (Salt in shaker)	R W DK	67
b. What is this taste? (Sugar in shaker)	R W DK	
c. What does this smell like? (Banana flavoring)	R W DK	
d. What does this smell like? (Onion flavoring)	R W DK	
15. Here's a pencil. Let me see you write your name. (Score for holding pencil correctly)	R W DK	68

	Instruction	Response	Related TEACH Item #
16a.	What is this word? (Cat)	R W DK	69
b.	What is this word? (Yes)	R W DK	
c.	What is this word? (Car)	R W DK	
d.	What is this word? (No)	R W DK	
17a.	Hold up your right hand.	R W DK	71
b.	Show me your left foot.	R W DK	
18a.	Who says "meow?" .... Now you make the same sound.	R W DK	74
b.	Who says "woof?" .... Now you make the same sound.	R W DK	
19.	What happens when you put water in the freezer?	R W DK	76
20a.	Write the letter "A."	R W DK	80
b.	Write the letter "O."	R W DK	
c.	Write the letter "X."	R W DK	
21a.	What color is this? (Red)	R W DK	82
b.	What color is this? (Yellow)	R W DK	
c.	What color is this? (Orange)	R W DK	
d.	What color is this? (Green)	R W DK	
22.	How many checkers are there? (8)	R W DK	91
23.	Let me see you hop on one foot.	R W DK	93
24.	Let me see you skip.	R W DK	95
25.	Tell me your favorite story. ______________________________		
26.	(Start with a row of 5 black and a row of 5 red checkers, with red and black checkers directly over each other. Say:) Look at the two rows of checkers. Now watch what I do. (Spread out row of red checkers) Are there as many red checkers as there are black, or are there more red checkers? (If child answers incorrectly, pair off red and black checkers again and say:) See, there is one red checker for every black checker. Do you see now that there are as many red checkers as black ones? Watch what I do. (Spread red checkers.) Now, are there as many red checkers as black checkers?	R W DK	99
27a.	What is this? (Green light on traffic board)	R W DK	37
b.	What should you do when you see this? (Green light)	R W DK	
c.	What is this? (Yellow light)	R W DK	
d.	What should you do when you see this?(Yellow light)	R W DK	
e.	What is this? (Red light)	R W DK	
f.	What should you do when you see this? (Red light)	R W DK	
g.	What is this? (Stop sign)	R W DK	
h.	What should you do when you see this? (Stop sign)	R W DK	

Now what does the light say? (Green) The light says "go," so it's time for you to go back to your class.